AFRL-AFOSR-VA-TR-2017-0099

Great Computational Intelligence in the Formal Sciences via Analogical Reasoning

Selmer Bringsjord
RENSSELAER POLYTECHNIC INST TROY NY
100 8TH STREET
TROY, NY 12180

05/08/2017
Final Report

Air Force Research Laboratory
AF Office Of Scientific Research (AFOSR)/RTA2

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE *(DD-MM-YYYY)*	2. REPORT TYPE	3. DATES COVERED *(From - To)*
08-05-2017	Final Performance	15 Oct 2011 to 31 Dec 2016

4. TITLE AND SUBTITLE
Great Computational Intelligence in the Formal Sciences via Analogical Reasoning

5a. CONTRACT NUMBER

5b. GRANT NUMBER
FA9550-12-1-0003

5c. PROGRAM ELEMENT NUMBER
61102F

6. AUTHOR(S)
Selmer Bringsjord

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)
RENSSELAER POLYTECHNIC INST TROY NY
100 8TH STREET
TROY, NY 12180 US

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
AF Office of Scientific Research
875 N. Randolph St. Room 3112
Arlington, VA 22203

10. SPONSOR/MONITOR'S ACRONYM(S)
AFRL/AFOSR RTA2

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
AFRL-AFOSR-VA-TR-2017-0099

12. DISTRIBUTION/AVAILABILITY STATEMENT
A DISTRIBUTION UNLIMITED: PB Public Release

13. SUPPLEMENTARY NOTES

14. ABSTRACT
We have delivered on our promise to develop and demonstrate a new level of great computational intelligence (GCI) in the formal sciences, via an unprecedented integration of analogical and deductive reasoning (analogico-deductive reasoning, or just `ADR') and associated techniques. The plan for the sequel is as follows. We begin by very brie
y reviewing our concept of distance
that originally gave rise to the adjective `great' in the GCI research program (x2). Next, we provide a bird's-eye perspective on the entire GCI paradigm and program, with help from a key \master' graphic (x3). (This perspective includes the fact that the work reported on herein included a 1-year extension of the original Phase-1 grant. Phase 1 set up the initial elements of the GCI paradigm, and applied them, to a degree, in certain domains. In light of this temporal context, the present report includes description of chief achievements made during the 1-year extension.) We then rapidly recount some of the highlights of Phase-1 accomplishments (x41).

15. SUBJECT TERMS
autonomy, computational systems

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			LAWTON, JAMES

Unclassified	Unclassified	Unclassified	UU	**PAGES**	**19b. TELEPHONE NUMBER** *(Include area code)* 703-696-5999

Final Report:
Great Computational Intelligence in the Formal Sciences via Analogical Reasoning

FA9550-12-1-0003

Selmer Bringsjord, PI • John Hummel, Co-PI • John Licato, Co-PI

043017NY

Contents

List of Figures

1 Introduction and Overview

We have delivered on our promise to develop and demonstrate a new level of **great computational intelligence** (GCI) in the formal sciences, via an unprecedented integration of analogical and deductive reasoning (**analogico-deductive reasoning**, or just 'ADR') and associated techniques.

The plan for the sequel is as follows. We begin by very briefly reviewing our concept of *distance* that originally gave rise to the adjective 'great' in the GCI research program (§2). Next, we provide a bird's-eye perspective on the entire GCI paradigm and program, with help from a key "master" graphic (§3). (This perspective includes the fact that the work reported on herein included a 1-year extension of the original Phase-1 grant. Phase 1 set up the initial elements of the GCI paradigm, and applied them, to a degree, in certain domains. In light of this temporal context, the present report includes description of chief achievements made during the 1-year extension.) We then rapidly recount some of the highlights of Phase-1 accomplishments (§4[1]).

2 Brief Recap: The Concept of Distance in GCI; Formal Structures Required

The launch of the GCI research program was in significant part based on a realistic assessment of the cognitive "distance" between the starting point of problem-solving undertaken by heralded AI systems, and the solutions produced by these systems. At launch, we specifically explained that such AI systems as Deep Blue and Watson, despite all the fanfare, only manage to travel a short distance from inputs to outputs (or — using the customary agent-centric language of AI (Russell & Norvig 2009) — from percepts to actions). GCI, by contrast, requires a capacity to, on the part of the AI involved, traverse a significant distance. We now briefly recount the nature of GCI and cognitive disstance as laid out at the launch of the GCI research program.

Deep Blue didn't have GCI; it was merely a demonstration that enormous computational firepower of a limited and inflexible type, when channeled with great human ingenuity, can, in completely specified Turing-solvable but timewise-demanding games of perfect information, yield human-level performance. This is certainly a mouthful, but didn't we know this all along? An original exemplar of a coNP-complete problem is that of deciding whether an arbitrary formula in the propositional calculus is provable. And we knew straightaway that machines can on this Turing-solvable but *very* timewise-demanding problem out-perform humans. So we knew long, long ago that chess really is just too easy, and hence Bringsjord's (1998) pointing this out was quite superfluous. Parallel comments could be fairly made about Watson, the *Jeopardy!*-winning AI system.[2] We launched the GCI research program to change the landscape, dramatically. When Phase 1 of the GCI program was inaugurated, irrational optimism about AI was steadfastly ignored, and this policy is still in place. Unfounded optimism about computational intelligence that doesn't seem to relate to rigorous theory and associated practice is still out there, but such sanguinity (e.g., that seen among those who believe that the so-called "Singularity" is near; e.g., Kurzweil 2006)

[1] Note that two PhD dissertations (Taylor & Licato at RPI) in computer science have been enabled by AFOSR support of the r&d carried out thus far. Both dissertations of course credit AFOSR, and are remarkable in their own right, but in the interests of space we omit discussion of them here, save for overlap between them and the papers published in the course of the project.

[2] For an analysis of Watson in the context of QA challenges that, if met, would be further toward GCI, see (Govindarajulu, Licato & Bringsjord 2014).

is not of concern to us because of our formal-scientific orientation. In short, people are free to pontificate that human-level AI is soon enough to arrive, but we politely ask to see the proofs supporting this optimism, and we politely ask to also see the human-level proofs and problem-solving exploits produced by these supposedly amazing machines, in support of their wondrous outputs. In the meantime, we intend to simply energetically work on developing the (GCI) resources for and demonstrations of human-level computing machines.

We seek GCI in the *formal sciences*, or in domains that can be, at least in significant part, formalized. We are concerned with computational intelligence in the realm of logic and mathematics, in fields and sub-fields based directly on them (e.g., reasoning and decision theory), and in informal domains that can be rigorized and hence brought into this realm, at least partially. In this general space, there is structured content in principle amenable to machine processing, and often results can be conclusively verified.

What does it take for some computational artifact to possess GCI? In general, for a computational artifact $\mathcal{C}$ to have GCI, we hold that it must produce a result ρ that is,

Significant by at least near-consensus among relevant humans, intrinsically significant;

Independent generated by a problem-solving run carried out to a high degree by $\mathcal{C}$ independent of human insight and assistance; and

Innovative where this problem-solving run begins from a starting point ι that is a "long distance" from ρ.

We shall assume that λ applied to a pair (ι, ρ) yields a distance δ; we therefore write

$$\lambda(\iota, \rho) = \delta.$$

To say that $\mathcal{C}$ produces ρ having started with ι, we write

$$\mathcal{C} : \iota \longrightarrow \rho.$$

We further assume that the general space of inputs is $\iota^\star$, and the general space of results $\rho^\star$. Under this notation, it can be informatively said that a good indicator of whether a result is significant is that the function f from $\iota^\star$ to $\rho^\star$ is Turing-*uncomputable*. Were this indicator promoted to an absolute requirement, which is quite tempting, the first property of GCI could plausibly be formalized via something like the following equation as a necessary condition for this property (significance) to be possessed.[3]

$$\mathcal{C} : \iota \longrightarrow \rho \text{ where the function } f : \iota^\star \longrightarrow \rho^\star \text{ is Turing-unsolvable.} \tag{1}$$

Our work in Phase 1 of the GCI research program has delivered GCI agents that satisfy the above conditions. But we are far from finished, as the remainder of the proposal makes clear.

[3]One must be careful here. Let h be a binary halting function taking as input the Gödel number n^M of a Turing machine M along with input m to that Turing machine. As is well-known, h is Turing-uncomputable. Yet there are individual Turing machines, accompanied by inputs to them, which can be instantly declared and proved to be either halters or non-halters as the case may be — by, say, Turing machine M^*. Therefore, by equation **??**, it would seem to follow that one of the necessary conditions for significance has been satisfied by M^*. What this example shows is that there can't be any "monkey business" going on with respect to what members of the range (= members of $\iota^\star$) the would-be GCI candidate is tackling. We leave aside how to formally regiment this requirement, but the intuitive idea should be clear, and there are many examples available to help characterize it.

For example, in the domain of mathematical logic, our new, ambitious targets now mark a clear instantiation of the characterization of GCI just given. But more importantly, inherent in new these new targets is a partial[4] metric for forming and measuring a fully formal concept of distance (δ): namely, length of minimal proof that the output is obtainable from the input, relative to some threshold $t \in \mathbb{N}$. (See below the central role that proof length plays even in the mere *exposition* of the two targets, one — the Gödel-Rosser Theorem — described in §**??**, and the other — Gödel's Speedup Theorem — in §**??**.) We do not yet have a precise threshold, but are working on it (see footnote 4).

A question regarding GCI and the concept of distance it includes remains:

- If length of proof is a major part of the concept of distance in play here, and hence a major part as well of the measurement of distance, what happens when the GCI agent is working in domains that don't always permit proofs?

For example, strategy on the adversarial side of the application of GCI, as shown clearly in Figure 3, can't always be such that that the outputs produced (e.g., recommendation of a course of action for blue-force actors) can be proved from relevant inputs and background information. The answer is actually straightforward: The concept of proof can simply be expanded to included the concept of *argument*, where the declarative information in arguments is still borne by formulae in formal languages, and the inferences, while now allowed to be non-deductive (analogical, inductive, abductive, etc.), must still conform to mechanically checkable inference schemata. Indeed, this answer aligns with clockwork precision with the transition of the GCI program into Phase 2, and specifically the completely new formalisms and mechanisms of analogico-*in*ductive reasoning (AIR) (see §5.3).

[4]The concept that completes the metric is the level of minimal **expressivity** required for the proof in question, given present science and technology. On this combined or two-part metric, based on length and expressivity, the "greatest" AI agent we have brought life, at least in part, is one to approach producing Goodstein's Theorem (GT), for the simple reason that current science and engineering doesn't allow a proof that can be expressed using only finite constructions. See (Govindarajulu, Licato & Bringsjord 2013). The present proposal isn't the place to articulate the two-part metric we sketch here; a paper is forthcoming, and the reader doubtless gets the basic idea.

3 High-Altitude View of GCI Paradigm and Program

The entire GCI research program can be profitably viewed from 30,000 feet via the "master graphic" shown in Figure 1. The master graphic hopefully instantly conveys what is perhaps our most significant discovery: the powerful cross-fertilization that occurs when developing and using GCI resources in different domains. For instance, early on in Phase 1, when we were trying to gradually grow the nascent resources that we knew would necessary down the road for obtaining Gödel's first incompleteness theorems from first principles, it occurred to us that it might make sense to test the nascent resources on an easier problem, not a tough, rarefied one in mathematical logic. So, we turned to problems that Piaget gave to children. And then what happened was that even though the resources at this point were embryonic, we met with success. This opened the door to aggressive cross-fertilization — and we have never looked back. Soon thereafter, we suspected that since a high-powered group of mathematical physicists in Hungary (at the Alfréd Rényi Institute of Mathematics) were working (in paper-and-pencil fashion) with many of the elements of mathematical logic to do physics, there was no reason why GCI resources wouldn't work in formal physics. Our suspicion turned out to be correct, and we developed a deep collaborative relationship with key researchers at the Rényi.

Returning to Figure 1, note that, time-wise, it summarizes Phase 1, which is done; and it depicts separately the 1-year extension, also done, and reported in the present document. The top box shows the chief components of the GCI "armamentarium." GCI is shown applied to two general clusters of domains: those falling under *formal science*, and those that are *adversarial*. (The concept of 'applied' here does *not* mean that we move beyond basic research, which should be, and at any rate will be, evident.) As the figure then shows, each cluster is broken out further, and progress in a trajectory for each domain is timelined, with each major milestone shown. Though this picture is a bird's-eye one, it is our record (and for that matter our guide, moving forward).

4 Work Accomplished in Phase 1 via GCI-ADR (Selected)

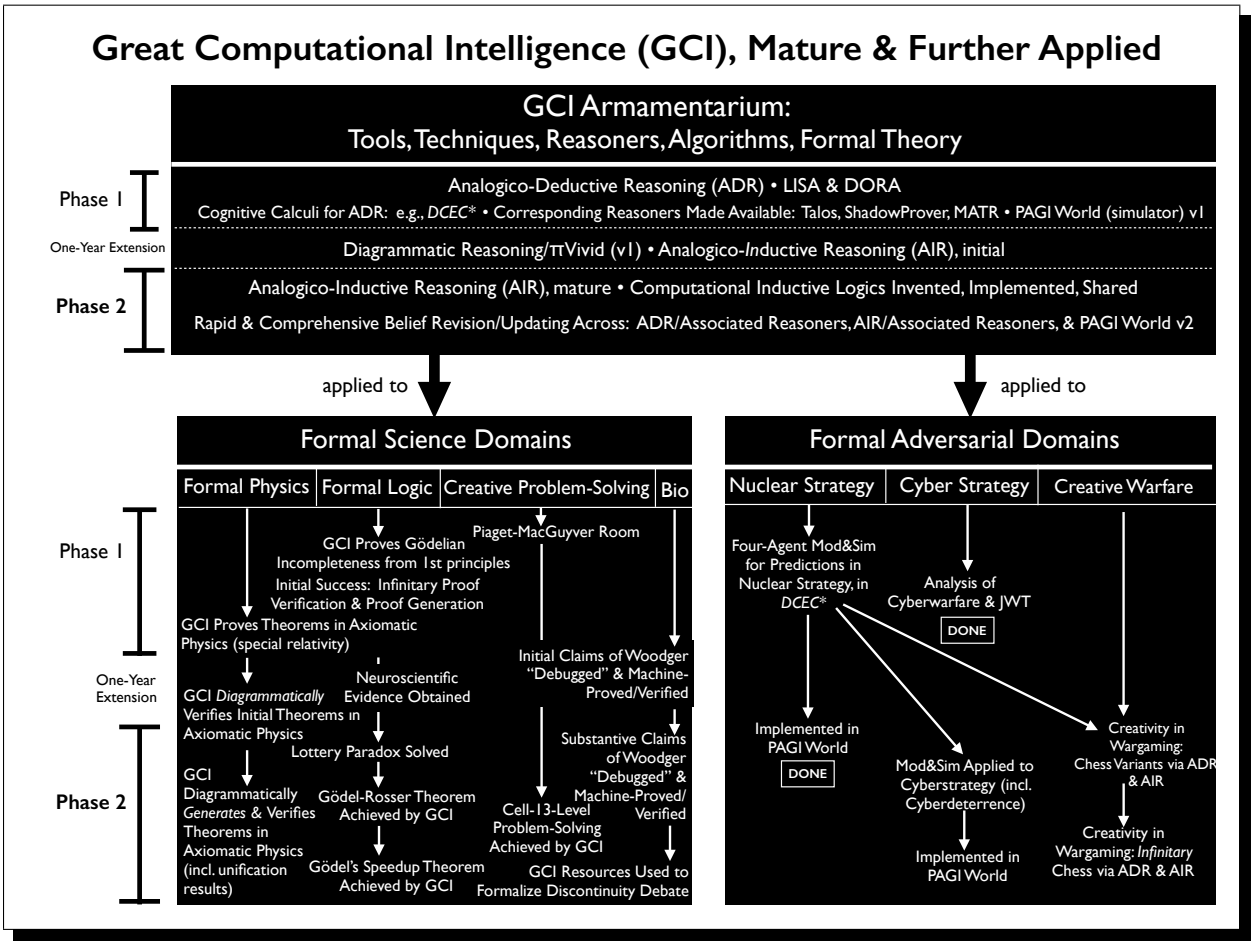
4.1 In Mathematical Logic

Analogico-deductive reasoning (ADR) is a type of reasoning used by humans in a variety of situations, producing many behaviors that are considered to be at the pinnacle of human intelligence. We describe our modeling of instances of ADR in the domain of mathematical logic (this section) and in creative human-level problem solving (§4.2).

Before explaining ADR in detail, note that analogical reasoning can often be useful in generation of hypotheses and theories to explain unfamiliar phenomena. For example, Holyoak et al. (2001) explain that the wave theory of sound, as it became better understood, was the basis for creating an analogy that described the wave theory of light. Such an analogical mapping would presumably be responsible for inferring the existence of a medium through which light would travel, just as sound needs air or something like it (indeed, the luminiferous aether was of course proposed to be this very medium). In contrast, Newton's particle theory of light would provide an analogical mapping that would not require a medium. Thus, we have two different analogical mappings; and each then suggests slightly different groups of hypotheses, members of which, in both cases, could in turn be tested with a combination of experimentation and deductive reasoning.

Analogico-deductive reasoning (ADR) is the intersection of hypothetico-deduction and analogical reasoning. In particular, we focus on a subset of ADR that involves the formation of hypotheses

Figure 1



h_i and the generation of results r_i which must be true if the h_i are true. Analogical reasoning can be responsible for either the formation of the hypotheses, or the conditionals of the form $h_i \rightarrow r_i$. Through further reasoning and experimentation in order to determine the validity of the r_i, the h_i can be either supported or rejected, ultimately by inference schemata that include such familiar, time-honored ones as *modus tollens* and *reductio ad absurdum*.

ADR is frequently used by individuals trying to understand new or unfamiliar concepts, as in the wave vs. particle example previously given. This is the case even in the highest levels of logical and mathematical reasoning, as when Kurt Gödel discovered and proved his incompleteness theorems. Gödel has been described as having a "line of thought [which] seems to move from conjecture to conjecture" (Wang 1995). In (Licato, Govindarajulu, Bringsjord, Pomeranz & Gittelson 2013), we use ADR to produce a complete and verifiably correct proof of Gödel's first incompleteness theorem **G1**, from first principles. Our source and target domains are **LP** (for the Liar Paradox) and **G1**, respectively. However, the transfer of knowledge between the two domains is not a trivial task: Whereas **G1** is completely formal and in a domain dealing with formal theories (of arithmetic) and well-formed-formulas, descriptions of **LP** are often semi-formal at best, and problems which claim to be equivalent to **LP** have involved objects as diverse as barbers in small towns, knights,

knaves, guards at bridges, and ancient Cretans. Therefore, an informal understanding of **LP** is to be used as no more than a *guide* to the discovery of **G1**. Much more creative thought is required to formulate concepts such as Gödel numbering or primitive-recursive functions.

LP, in its simplest form, is a self-referential statement that at first glance seems cute and unproblematic, but upon closer scrutiny leads to an apparent contradiction. "This statement is false" is a typical version; the problem appears when one tries to determine if the statement is actually true or not. If true, it's false; if false, it's true. Thus, a contradiction appears if we assume that all such statements must have one and exactly one truth value. The Gödel statement, which may be informally stated as "This statement is not provably true," faces a similar problem. If it's provably true, it's false, and thus it can't be provably true; if it's provably false, then the statement is provably true, and this implies that the reasoning system in play here is inconsistent.

Figure 2 shows the proof of **G1**, generated through ADR. See (Licato et al. 2013) for details.

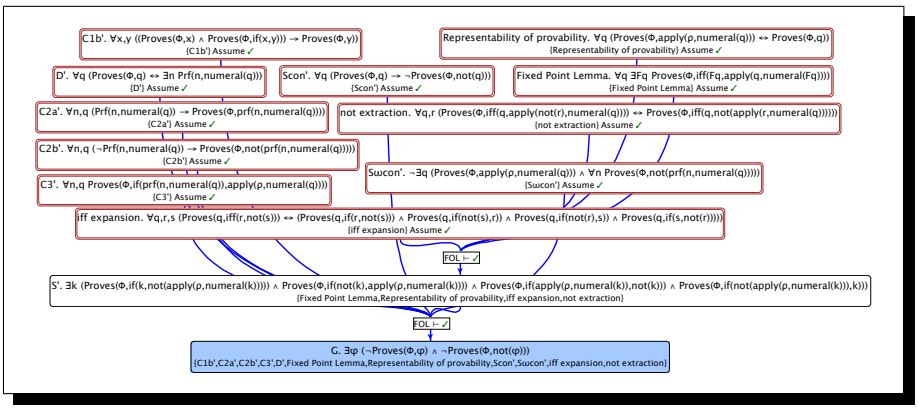


Figure 2: Partial Proof of **G1**, Pictured in Slate

Another proof whose discovery can (at least partially) be explained through ADR is Goodstein's Theorem:

[**Goodstein's Theorem**] For all natural numbers, the Goodstein sequence reaches zero after a finite number of steps.

The Goodstein sequence is the sequence obtained by repeated iterations of a growth function, pictured in Figure 3. Goodstein's theorem is of interest because it is not provable in Peano Arithmetic (PA) (or any equivalent theory of arithmetic); this makes it an instance of Gödelian incompleteness. All known proofs of Goodstein's Theorem use infinitary constructs one way or another. The proofs either require infinite sets (beyond finitary arithmetic theories such as PA), or the proofs require non-finitary rules such as the $\omega-$rule, an inference rule with an infinite number of premises. However, because of its infinitary nature, the $\omega-$rule is unsuitable for implementation in a standard, non-hypercomputational computer. The advantage of the variant of the rule shown below is that PA in a system which has this rule is negation-*complete*; that is, for every ϕ in the appropriate language, either $\mathsf{PA} \vdash_\omega \phi$ or $\mathsf{PA} \vdash_\omega \neg\phi$.

A restricted form of the $\omega-$rule is a finite form which still preserves negation-completeness for PA. The disadvantage is that proof-verification and discovery both fail to be even semi-decidable (as proved in Govindarajulu & Bringsjord 2014). Even this restricted form is beyond trivial machine

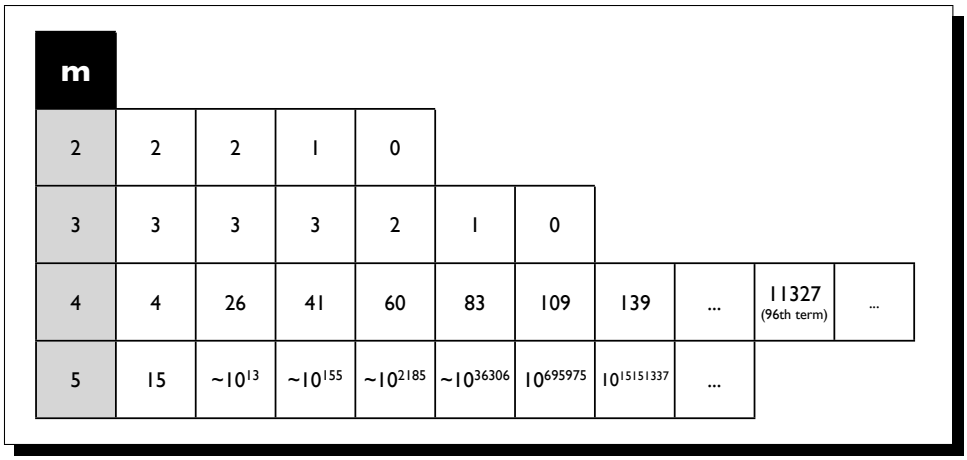


Figure 3: The Goodstein Sequence's Explosive Growth

implementation, as, in the general case, a proof-verification system that handles the rule should be able to check in all possible cases if the program supplied halts with the correct proof.

(Govindarajulu et al. 2013) produces a partial proof of Goodstein's Theorem using ADR, by using theorems and proofs from a source domain containing finite concepts (more familiar to the average mathematician) and a target domain containing the infinitary concepts needed. A reasoner can then produce a proof of crucial parts of the proof of Goodstein's Theorem. The output of the target domain is pictured in Figure 4.

4.2 In Creative Human-Level Problem-Solving

Although ADR might explain some of the most impressive feats of advanced logico-mathematical reasoning, such as finding the proofs of **G1** and Goodstein's Theorem, it turns out that ADR is also useful in modeling the reasoning done by children on Piagetian tasks.

In (Bringsjord & Licato 2012), ADR is applied to a Piagetian task known as the *magnet task*, in which a circular board is placed in front of children. The board contains a metal rod anchored to the center, so that it can be spun and ultimately rest pointing to a pair of boxes placed along the circumference of the board. Each of the pairs of boxes have a simple shape drawn on them. Unbeknownst to the child, one pair of boxes contains hidden magnets, so that the rotating metal rod is guaranteed to settle in a position that points to the boxes containing the magnets. The child is then asked to explain why the rod behaves as it does.

One child, referred to as 'Gou', does from the start suspect that magnets are responsible — but quickly abandons this hypothesis in favor of the one claiming that the weight of the objects is what leads the needle to repeatedly stop on the stars. The experimenter then asks Gou what he would have to do in order to "prove that it isn't the weight," to which Gou responds by carrying out a series of small experiments designed to prove that weight isn't responsible for the bar's stopping. One of these experiments involves removing the star and diamond boxes, and checking to see if the bar still stops on the heaviest of the remaining boxes. Predictably (given *our* understanding of the background mechanisms), it does not; this provides Gou with empirical evidence that weight is not causally responsible for the bar's stopping as it invariably does (although he continues to subsequently perform small experiments to further verify that weight is not responsible). We model

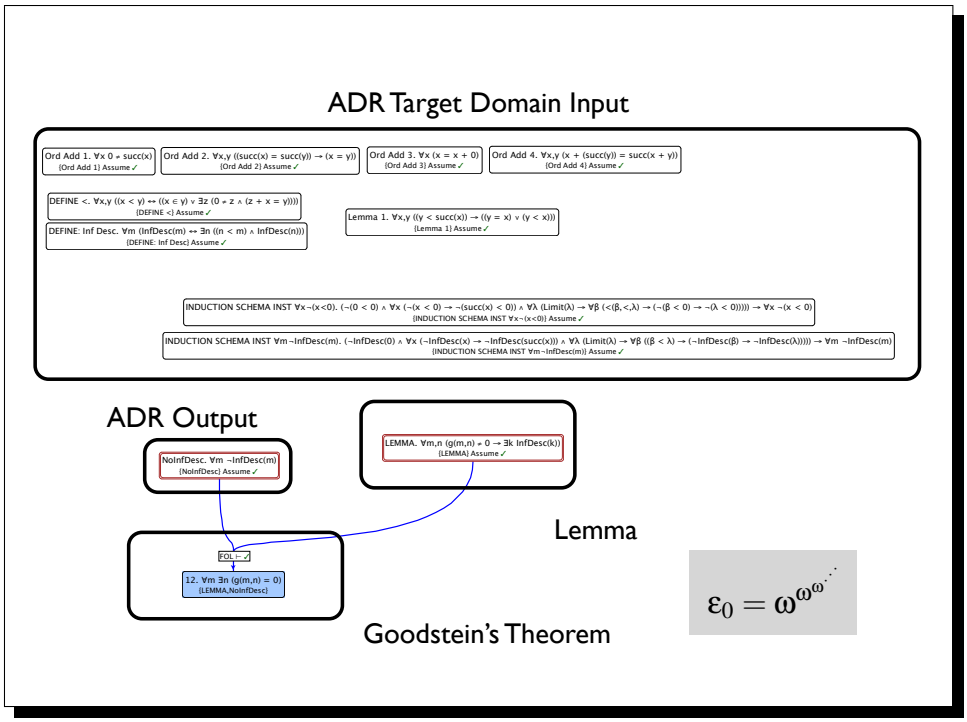


Figure 4: Output of an ADR process applied to proving Goodstein's Theorem; for details see (Govindarajulu et al., 2013).

the analogical inference used by Gou (Figure 5) by using the computational model of analogy LISA (Hummel & Holyoak 1997, Hummel & Holyoak 2003).

Another Piagetian task, the *Balance Beam Task* (BBT), asks subjects to explain the behavior of a balance beam where weights can be hung at different distances from the center. Again, we model the reasoning process of a child working his way through this task using LISA (Licato, Bringsjord & Hummel 2012), and the resulting inference is pictured in Figure 6.

Tasks such as the magnet task and BBT may be part of a collection of psychometric tasks to detect artificial *general* intelligence (AGI). There may be countless other PAGI (psychometric AGI) tasks, and in order to administer as many of them as possible in an efficient and inexpensive way, we make use of the physically-realistic simulation environment PAGI World (Licato & Bringsjord 2015). PAGI World allows for tasks to be quickly deployed and to be administered easily to virtually any artificial-intelligence script or cognitive system, so long as that script or system can communicate through TCP/IP. Figure 7 shows an agent in PAGI World solving another Piagetian task, described by Licato, Marton, Dong, Sun & Bringsjord (2015).

4.3 In Axiomatic Physics

While there has been considerable progress in proof verification in the formal sciences, for instance the Mizar project and the four-color theorem, now machine verified, there has been scant such work carried out in the realm of the natural sciences. The delay in the case of the natural sciences can be attributed to both a lack of formal analysis of the informal theories in such sciences, probably due to an inclination toward informality and empiricism on the part of nearly all of the relevant

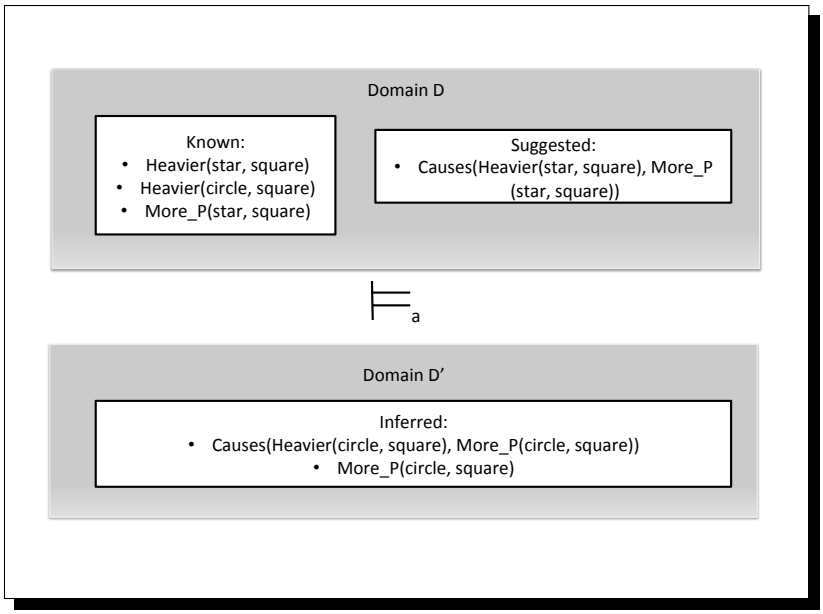


Figure 5: Propositional Knowledge Used in LISA for the Gou Example. Domain D' is inferred from D using LISA's analogical inferencing capability. Propositions representing semantic connections are not pictured here.

scientists. In (Govindarajalulu, Bringsjord & Taylor 2015), we gave a compressed report on our building upon these formal theories using logic-based AI in order to achieve, in relativity, both machine proof-discovery and proof-verification, for theorems previously established by humans. Our report was intended to serve as a springboard to machine-produced results in the future that have not been obtained by humans.

Our work could be compared with the recent work (Stannett & Németi 2014), in which a proof for Theorem NFTL, discussed later, is formally verified. While our two projects start at the same point, our goals are different. Stannett and Németi strive to formally verify whether the computational power of devices depend on the nature of the spacetime considered. Our goal, in keeping with the GCI paradigm, is to enable and explore tools for automatic verification and discovery of formal knowledge in the natural sciences.

Inspired by how an insightful user of a monolithic theorem prover can efficaciously break up problems, we built a simple inference system **semi** that models semi-automated theorem proving with a sound and complete monolithic theorem prover. For proof verification, we use a system built upon a denotational proof language (Arkoudas 2000). We then combined both proof verification and proof discovery in an easy-to-use environment, Slate (Bringsjord, Taylor, Shilliday, Clark & Arkoudas 2008), which used the automated theorem prover SNARK [5].

We then reported machine-generated and machine-verifiable proofs of theorems in (Andréka, Madarász & Németi 2007). Our initial result focused on the special theory of relativity, which was formalized in that work. We have obtained, in Slate, a machine-verified proof, as well as a machine-generated proof, of a theorem whose givens include lemmas from field theory. The theorem, dubbed 'NEAT,' states that for no inertial observer are the events at one point the same as the events at

[5] http://www.ai.sri.com/ stickel/snark.html

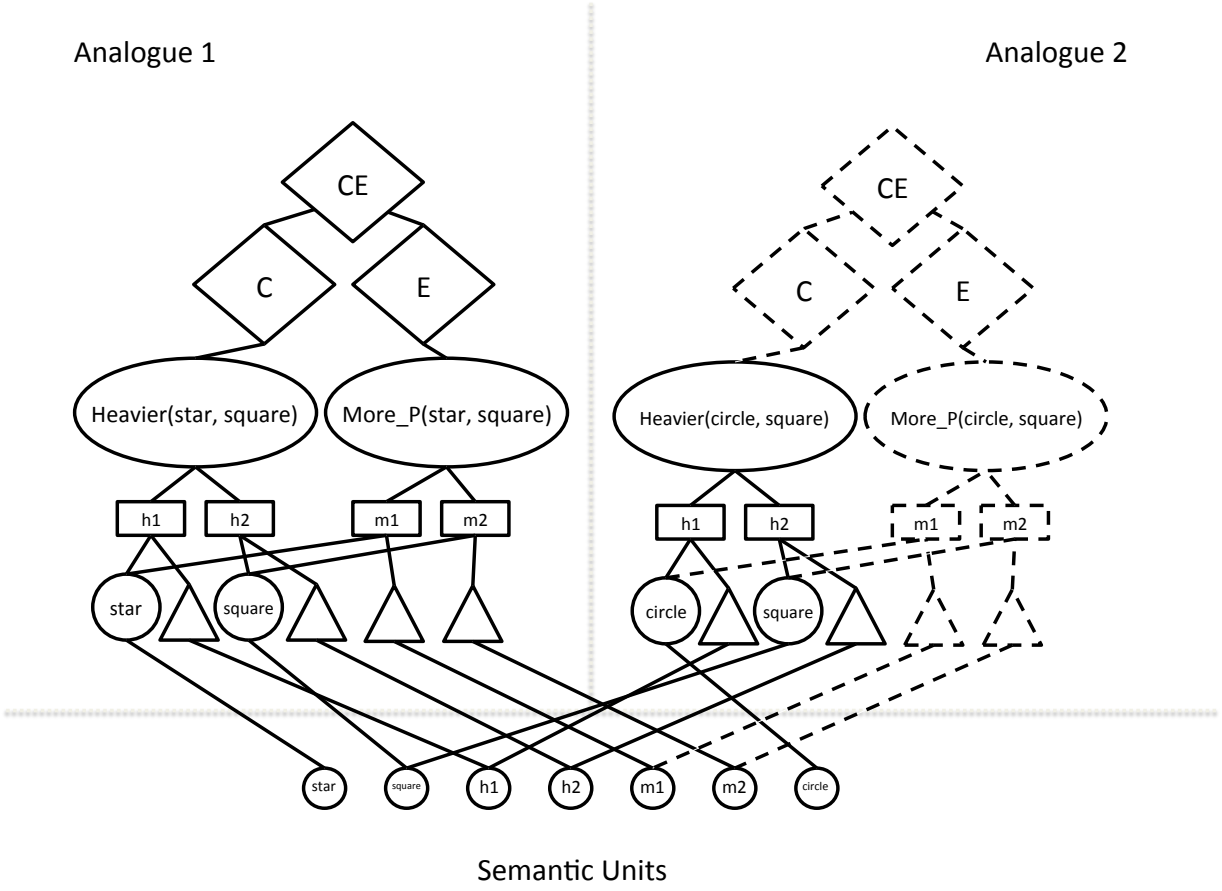


Figure 6: The Analogs Used in Our Model. For simplicity, not all individual semantic units, propositions, or other units are pictured here. Dotted lines represent units that were recruited as a result of analogical inference.

another point. The full proof using natural deduction in Slate is shown in Figure 8.

We have also obtained a formal proof of Theorem 2.1 in (Andréka et al. 2007), and, in accordance with our aims, have verified the proof. This theorem states that no inertial observer m can observe another inertial observer k travelling faster than the speed of light. The full proof is too big to present in the present document.[6]

4.4 In Axiomatic Nuclear Strategy

$\mathcal{DCEC}^*$ (deontic cognitive event calculus) is a family of *multi-sorted quantified modal logics*[7] that has a well-defined syntax and a proof calculus. The syntax of the language of a dialect of $\mathcal{DCEC}^*$

[6]The code for the whole proof in the semi-automated layer can be obtained at this url: https://s3.amazonaws.com/SyntheseSubmission/logicphysics.zip. If there is a problem obtaining, please contact S. Bringsjord.

[7]Manzano (1996) covers muli-sorted first-order logic (MSL). Details as to how a reduction of intensional logic to MSL so that automated theorem proving based in MSL can be harnessed is provided in (Arkoudas & Bringsjord 2009*a*).

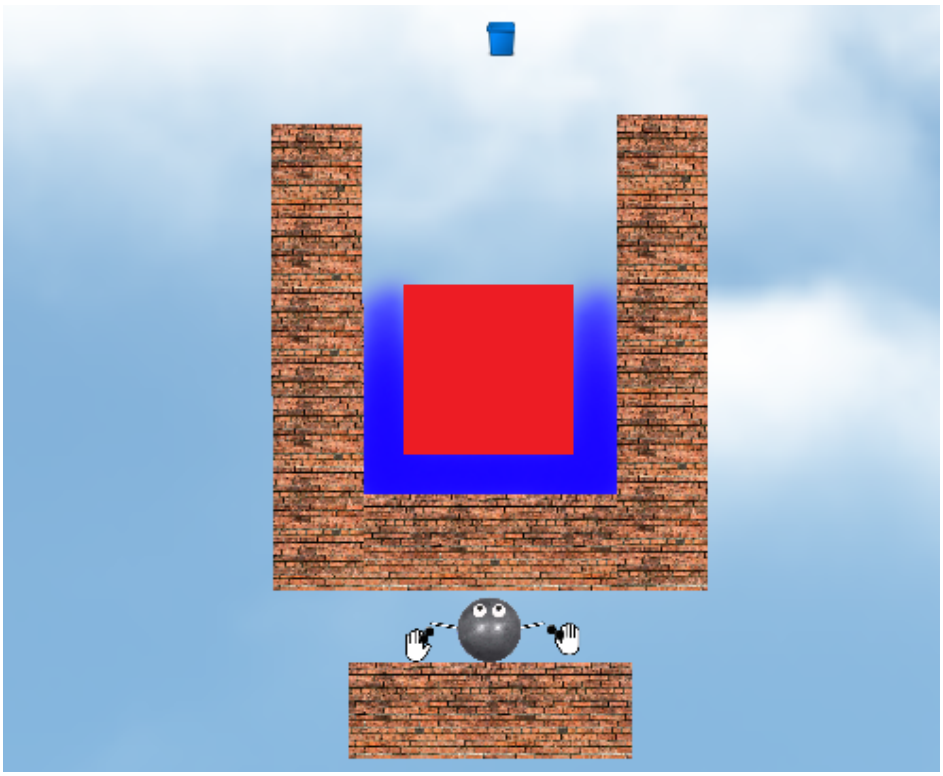


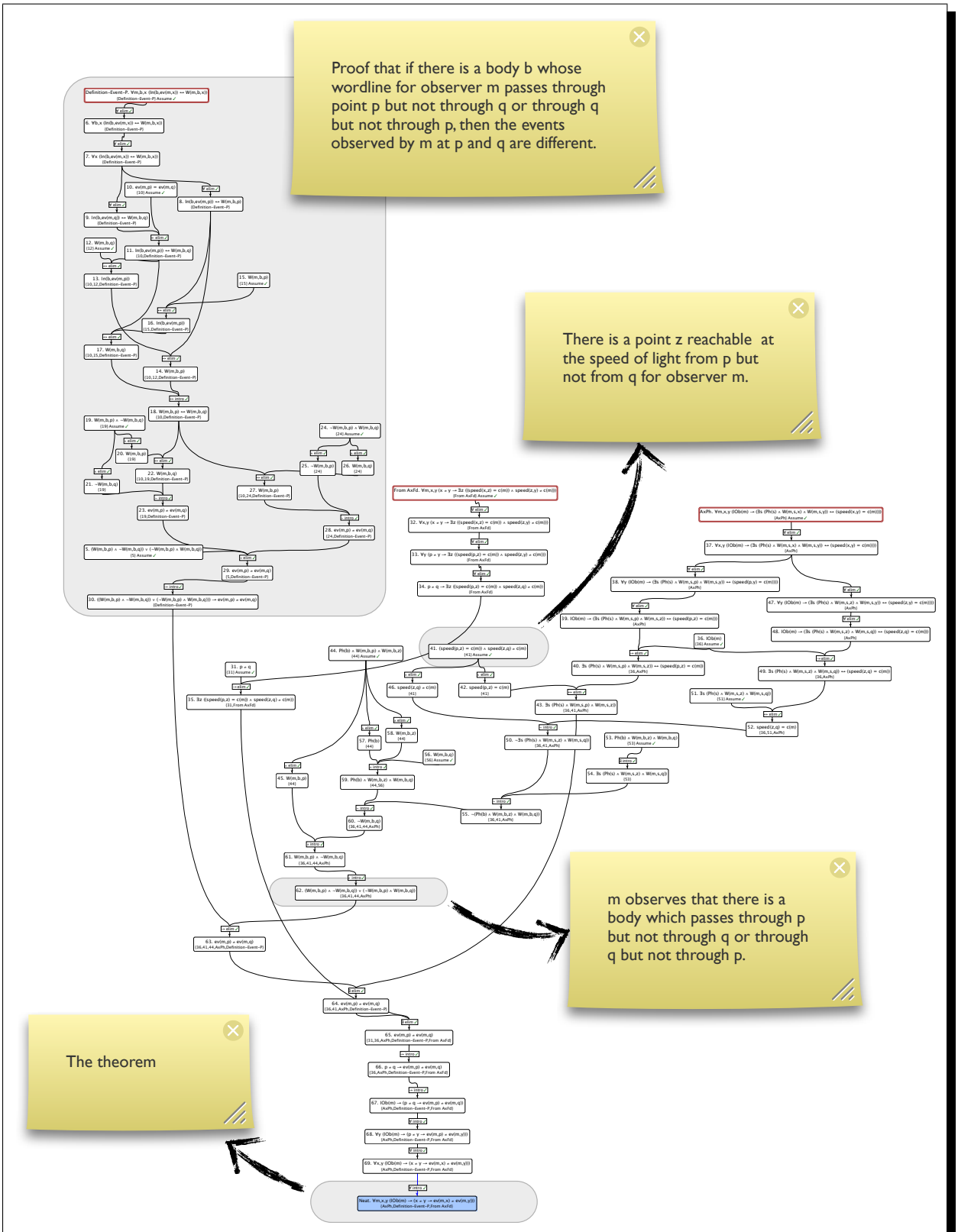
Figure 7: An Agent Solving a Piagetian Task in PAGI World

and the inference schemata for this dialect's proof calculus are shown in Figure 9. $\mathcal{DCEC}^*$ syntax includes a system of sorts S, a signature f, a grammar for terms t, and a grammar for sentences ϕ; these are shown on the left half of the figure. The proof calculus is based on natural deduction (Jaśkowski 1934), and includes all the introduction and elimination schemata for first-order logic, as well as those for the intensional operators; the master list of schemata are shown on the right half of the figure.

Bringsjord et al. (2014) use an early dialect of $\mathcal{DCEC}^*$ to consider increasingly complex models of nuclear strategy. The first two models are simple, in that their structure is fixed and the only possibility of variation is through adjustment of parameter values. Specifically, there is no provision for incorporating deliberative mind-reading[8] in these two models. The third model builds upon the first two and uses a dialect of $\mathcal{DCEC}^*$ to specify the model, and accordingly has enough expressive power to capture mindreading by the players involved. In addition to mindreading, the third model can also capture any arbitrary scenario that could be of relevance. For example, the first two models are agnostic on whether communication between the U.S. and Israel could be monitored by Russia for Iran. If we want to look at the effects of Russia monitoring such communication, we could, in principle, supply a statement of this fact and other relevant information in the form of a set of statements Γ_{ND} to a semi-automated implementation of our cognitive calculus, and ask the system questions ϕ that we might be interested in (where ϕ contains information about the relevant deterrence scenario). We argue that the cognitive calculus is expressive enough to model

[8]In short, deliberative reasoning about the beliefs, knowledge, plans, goals, intention, language, emotions, ... of agents.

Figure 8: Full proof of Theorem NEAT



deliberative mindreading. This entails that the formal calculus contain, at a minimum, syntax for expressing intensional operators like *knows*, *believes*, *ought*, and for expressing time, change, events, and actions.

It's particularly important to realize that in modeling nuclear strategy, we are ultimately interested in answering the following question via simulation (and subsequent verification thereof):

$$\Gamma_{ND} \vdash_{\mathcal{DCEC}^*} happens(actions(iran, attack(israel)), T)?$$

In (Bringsjord et al. 2014), the first two models presented have assumed that the agents are ideal mechanical information-processors, in that they adhere to the model perfectly without any variation, and have access to all the relevant information, which makes mindreading unnecessary. The models can be made more realistic by introducing cognitive agents who do not have access to perfect and complete information about what the other agents are thinking; this makes mindreading essential. These agents are agents modeled in $\mathcal{DCEC}^*$. They perform actions according to the dynamic model developed above, but the costs involved and the actions of other agents are not known in advance: the agents have beliefs about the costs. Agents can manipulate these beliefs by public communication of threats and other actions. In short, mindreading becomes essential, and makes the model move toward what is real-world.

The relevant $\mathcal{DCEC}^*$ agent symbols corresponding to the states are: $\{us, iran, russia, israel\}$. The U.S. has ability to perform an action of the following type: $deter : Numeric \rightarrow ActionType$, which means an action of a specific intensity to deter Iran performed by the U.S. If $\overline{c}$ is a $\mathcal{DCEC}^*$ term that denotes the natural number c, we can take $deter(\overline{0})$ to be a moderate economic action, $deter(\overline{1})$ to be a severe economic sanction, $deter(\overline{2})$ to mean limited military action, and so on. For convenience, we use the sloppier $deter(c)$ hereafter in the present exposition. Iran can enhance its capability at a certain cost $enhance : ActionType$. Iran can also attack Israel: $attack : Agent \rightarrow ActionType$. The cost of enhancing Iran's nuclear capability at any time is given by $cost : Moment \rightarrow Numeric$. In addition to the above explicit actions, the participants can engage in communication at any time t of declarative information ϕ between themselves, modeled by the $says(a, b, t, \phi)$ operator; or public communication, modeled by $says(a, t, \phi)$.[9] The fluent *capable* denotes whether Iran's capacity is susceptible of being morphed into attack capability against Israel by time T. The fluent *destroyed* denotes whether Israel has been destroyed.

Given that Iran has enough cognitive power for simple utilitarian calculations, we can derive (using a proof explained in Bringsjord et al. (2014)):

$$\forall t : Moment\ t < T \Rightarrow \neg \mathbf{B}(iran, t, cost_{sum}(t+1, T) > \Delta - deter_{sum}(t+1, T))$$

From the above formula, we can obtain a simple first-order derivation that shows that deterrence fails; an automatic proof of it is given in Figure 10. Figure 11 shows the fully machine-generated proof based on resolution and paramodulation in the SNARK automated theorem prover (see Stickel et al. (1994)).

[9]Note that the two forms of *says* are in fact a different kind of syntactic object, but are represented by a similar letter for convenience. We have not included a group-communication operator and we acknowledge that we are ignoring intricate subtleties involved in communication, since a simple dialect of $\mathcal{DCEC}^*$ without NLP capability has been used.

Syntax

$$S ::= \text{Object} \mid \text{Agent} \mid \text{Self} \sqsubset \text{Agent} \mid \text{ActionType} \mid \text{Action} \sqsubseteq \text{Event} \mid \text{Moment} \mid \text{Boolean} \mid \text{Fluent} \mid \text{Numeric}$$

$$f ::= \begin{array}{l} \mathit{action} : \text{Agent} \times \text{ActionType} \to \text{Action} \\ \mathit{initially} : \text{Fluent} \to \text{Boolean} \\ \mathit{holds} : \text{Fluent} \times \text{Moment} \to \text{Boolean} \\ \mathit{happens} : \text{Event} \times \text{Moment} \to \text{Boolean} \\ \mathit{clipped} : \text{Moment} \times \text{Fluent} \times \text{Moment} \to \mathit{Boolean} \\ \mathit{initiates} : \text{Event} \times \text{Fluent} \times \text{Moment} \to \text{Boolean} \\ \mathit{terminates} : \text{Event} \times \text{Fluent} \times \text{Moment} \to \text{Boolean} \\ \mathit{prior} : \text{Moment} \times \text{Moment} \to \text{Boolean} \\ \mathit{interval} : \text{Moment} \times \text{Boolean} \\ * : \text{Agent} \to \text{Self} \\ \mathit{payoff} : \text{Agent} \times \text{ActionType} \times \text{Moment} \to \text{Numeric} \end{array}$$

$$t ::= x : S \mid c : S \mid f(t_1, \ldots, t_n)$$

$$\phi ::= \begin{array}{l} t : \text{Boolean} \mid \neg\phi \mid \phi \wedge \psi \mid \phi \vee \psi \mid \forall x : S.\ \phi \mid \exists x : S.\ \phi \\ \mathbf{P}(a,t,\phi) \mid \mathbf{K}(a,t,\phi) \mid \mathbf{C}(t,\phi) \mid \mathbf{S}(a,b,t,\phi) \mid \mathbf{S}(a,t,\phi) \\ \mathbf{B}(a,t,\phi) \mid \mathbf{D}(a,t,\mathit{holds}(f,t')) \mid \mathbf{I}(a,t,\mathit{happens}(\mathit{action}(a^*,\alpha),t')) \\ \mathbf{O}(a,t,\phi,\mathit{happens}(\mathit{action}(a^*,\alpha),t')) \end{array}$$

Rules of Inference

$$\frac{}{\mathbf{C}(t,\mathbf{P}(a,t,\phi) \to \mathbf{K}(a,t,\phi))}\ [R_1] \qquad \frac{}{\mathbf{C}(t,\mathbf{K}(a,t,\phi) \to \mathbf{B}(a,t,\phi))}\ [R_2]$$

$$\frac{\mathbf{C}(t,\phi)\ t \le t_1 \ldots t \le t_n}{\mathbf{K}(a_1,t_1,\ldots \mathbf{K}(a_n,t_n,\phi)\ldots)}\ [R_3] \qquad \frac{\mathbf{K}(a,t,\phi)}{\phi}\ [R_4]$$

$$\frac{}{\mathbf{C}(t,\mathbf{K}(a,t_1,\phi_1 \to \phi_2) \to \mathbf{K}(a,t_2,\phi_1) \to \mathbf{K}(a,t_3,\phi_3))}\ [R_5]$$

$$\frac{}{\mathbf{C}(t,\mathbf{B}(a,t_1,\phi_1 \to \phi_2) \to \mathbf{B}(a,t_2,\phi_1) \to \mathbf{B}(a,t_3,\phi_3))}\ [R_6]$$

$$\frac{}{\mathbf{C}(t,\mathbf{C}(t_1,\phi_1 \to \phi_2) \to \mathbf{C}(t_2,\phi_1) \to \mathbf{C}(t_3,\phi_3))}\ [R_7]$$

$$\frac{}{\mathbf{C}(t,\forall x.\ \phi \to \phi[x \mapsto t])}\ [R_8] \qquad \frac{}{\mathbf{C}(t,\phi_1 \leftrightarrow \phi_2 \to \neg\phi_2 \to \neg\phi_1)}\ [R_9]$$

$$\frac{}{\mathbf{C}(t,[\phi_1 \wedge \ldots \wedge \phi_n \to \phi] \to [\phi_1 \to \ldots \to \phi_n \to \psi])}\ [R_{10}]$$

$$\frac{\mathbf{B}(a,t,\phi)\ \ \mathbf{B}(a,t,\phi \to \psi)}{\mathbf{B}(a,t,\psi)}\ [R_{11a}] \qquad \frac{\mathbf{B}(a,t,\phi)\ \ \mathbf{B}(a,t,\psi)}{\mathbf{B}(a,t,\psi \wedge \phi)}\ [R_{11b}]$$

$$\frac{\mathbf{S}(s,h,t,\phi)}{\mathbf{B}(h,t,\mathbf{B}(s,t,\phi))}\ [R_{12}]$$

$$\frac{\mathbf{I}(a,t,\mathit{happens}(\mathit{action}(a^*,\alpha),t'))}{\mathbf{P}(a,t,\mathit{happens}(\mathit{action}(a^*,\alpha),t))}\ [R_{13}]$$

$$\frac{\mathbf{B}(a,t,\phi)\ \ \mathbf{B}(a,t,\mathbf{O}(a^*,t,\phi,\mathit{happens}(\mathit{action}(a^*,\alpha),t')))\ \ \mathbf{O}(a,t,\phi,\mathit{happens}(\mathit{action}(a^*,\alpha),t'))}{\mathbf{K}(a,t,\mathbf{I}(a^*,t,\mathit{happens}(\mathit{action}(a^*,\alpha),t')))}\ [R_{14}]$$

$$\frac{\phi \leftrightarrow \psi}{\mathbf{O}(a,t,\phi,\gamma) \leftrightarrow \mathbf{O}(a,t,\psi,\gamma)}\ [R_{15}]$$

Figure 9: A Dialect of the Deontic Cognitive Event Calculus

Figure 10: Slate Proof that Deterrence Will Fail Under Some Circumstances

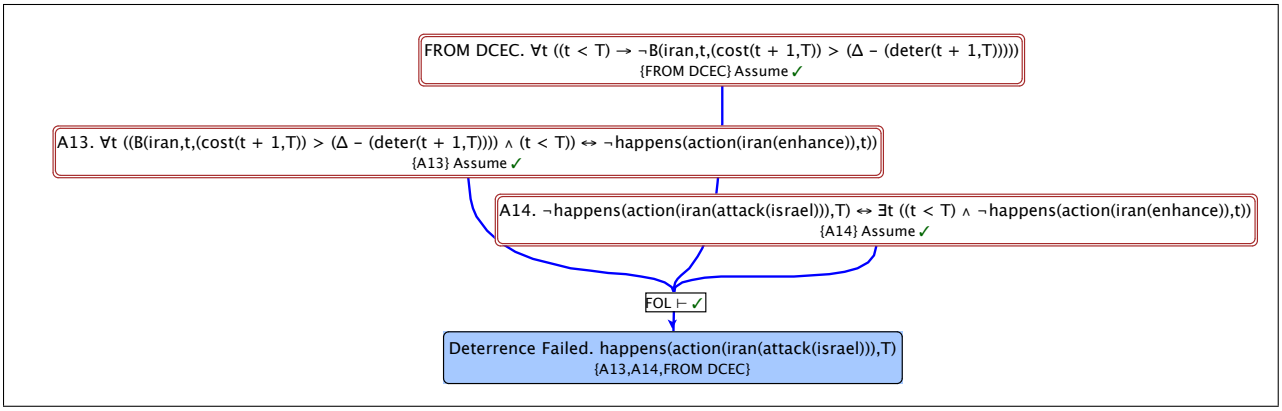
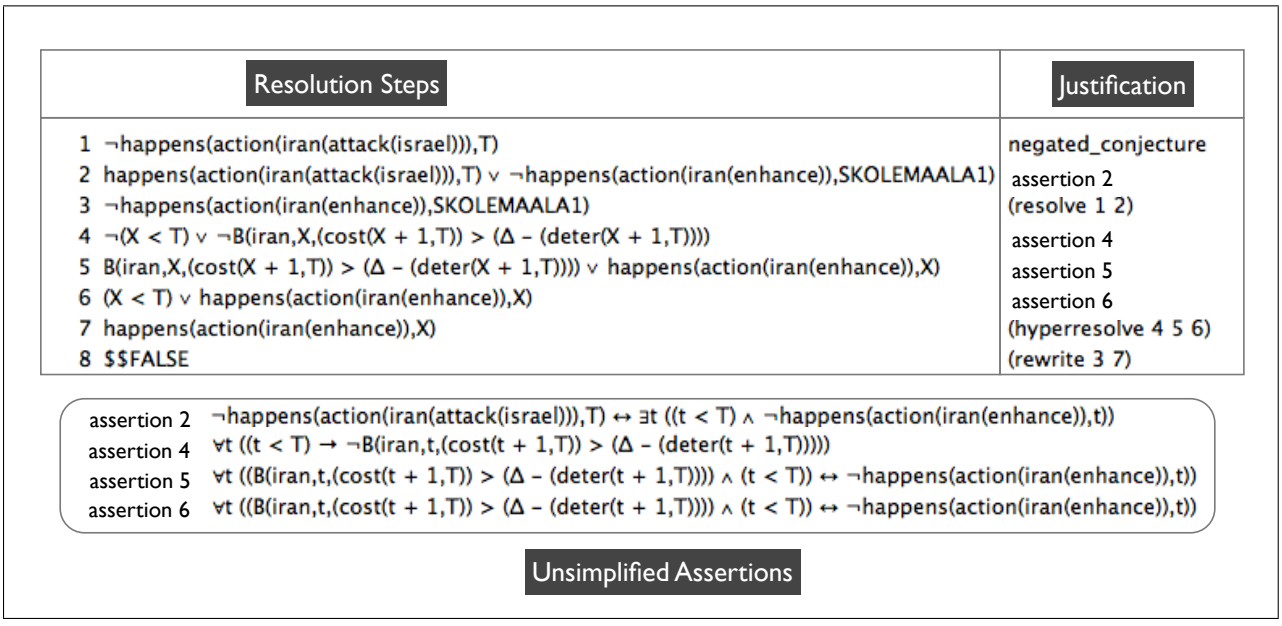


Figure 11: Internal SNARK Proof (Automatically Generated)



Resolution Steps	Justification
1 ¬happens(action(iran(attack(israel))),T)	negated_conjecture
2 happens(action(iran(attack(israel))),T) ∨ ¬happens(action(iran(enhance)),SKOLEMAALA1)	assertion 2
3 ¬happens(action(iran(enhance)),SKOLEMAALA1)	(resolve 1 2)
4 ¬(X < T) ∨ ¬B(iran,X,(cost(X + 1,T)) > (Δ − (deter(X + 1,T))))	assertion 4
5 B(iran,X,(cost(X + 1,T)) > (Δ − (deter(X + 1,T)))) ∨ happens(action(iran(enhance)),X)	assertion 5
6 (X < T) ∨ happens(action(iran(enhance)),X)	assertion 6
7 happens(action(iran(enhance)),X)	(hyperresolve 4 5 6)
8 $$FALSE	(rewrite 3 7)

assertion 2	¬happens(action(iran(attack(israel))),T) ↔ ∃t ((t < T) ∧ ¬happens(action(iran(enhance)),t))
assertion 4	∀t ((t < T) → ¬B(iran,t,(cost(t + 1,T)) > (Δ − (deter(t + 1,T)))))
assertion 5	∀t ((B(iran,t,(cost(t + 1,T)) > (Δ − (deter(t + 1,T)))) ∧ (t < T)) ↔ ¬happens(action(iran(enhance)),t))
assertion 6	∀t ((B(iran,t,(cost(t + 1,T)) > (Δ − (deter(t + 1,T)))) ∧ (t < T)) ↔ ¬happens(action(iran(enhance)),t))

Unsimplified Assertions

5 Report on Progress During 1-Year Extension

We now report on chief achievements made during the 1-year extension.

5.1 In *Diagrammatic* Reasoning in Formal Physics

Our prior work in axiomatic physics, like almost all proof-oriented work in the formal sciences, was homogeneously linguistic in nature: the proofs in question are based exclusively on formal languages; diagrams, pictures, images, etc. are nowhere to be seen. Yet undeniably mathematical physicists routinely employ (informal) visual and diagrammatic reasoning in their proofs, and especially so in the theories of relativity. A formal system leveraging both visual and symbolic reasoning enables heterogeneous proofs that are (i) not only more readable, intuitive, and consistent with scientific practice, but also (ii) simpler (in a formal sense), and therefore potentially easier for machines to discover on their own. During the 1-year-extension period, began investigating precisely such a system, one built directly atop Vivid (Arkoudas & Bringsjord 2009*b*), a heterogeneous logicist framework in turn built atop denotational proof languages. We employed the system to move closer to a formal, semi-automated proof of Theorem NEAT (recall above: 4.3) that is at once both linguistic and diagrammatic.

5.1.1 Progress During Current 1-Year Extension

We now give a sustained example (extracted from recent work in the 1-year extension) in order to partially show and thereby explain the extension of our work into the realm of proofs in formal physics that are partly visual/diagrammatic in nature.

This section concerns Theorem 2.2 from "Logic of Space-time and Relativity Theory" (Andréka et al. 2007), which states that no observer observes the same event at two different space-time locations in models of the field and light axioms of special relativity. We abbreviate this, as before, as 'Theorem NEAT' (No Event At Two Places). An informal proof is supplied in English; the proof uses geometric constructs in its reasoning. We now describe how this proof may be formulated in a formal heterogeneous (i.e., linguistic/sentential-and-diagrammatic) manner, and represented in a Vivid language for automatic verification.

> **Argument**: Consider a 2-dimensional plane with two axes, the horizontal axis representing the spatial dimension and the vertical axis representing the temporal dimension. Let there be an inertial observer m at the origin. Consider two distinct points x and y on the plane.
>
> From the field axioms, it follows that through any point x, there will be two lines of slope 1, where the speed of light $c = 1$.
>
> We now summarize the process of formulating a diagrammatic representation in Vivid, and provide such a formulation. To specify a formula instance in Vivid, three steps are necessary.
>
> First, we must provide an Attribute Structure A containing a collection of attributes and the set of their possible values, along with a set of computable relations $\mathcal{R}$ involving those attributes. We choose the attributes `position`, `slope`, and `line_positions` with ranges $\mathbb{R}^n$, $[0, 1]$, and $\mathbb{R}^{n^*}$ (where $*$ is taken to mean the set of all lists of the elements of $\mathbb{R}^n$), respectively. Furthermore, we choose the relations $R_1 \subseteq \mathbb{R}^n \times \mathbb{R}^{n^*}$ defined as $R_1(p_1, [p'_1, ..., p'_k]) \Leftrightarrow p_1 \in [p'_1, ..., p'_k]$, $R_2 \subseteq \mathbb{R}^n \times \mathbb{R}^n$ defined as $R_2(p_1, p_2) \Leftrightarrow p_1 = p_2$, and $R_3 \subseteq [0, 1]$ defined as $R_3(s) \Leftrightarrow s = 1$. Then, in Vivid's native notation, our attribute structure is $A = (\{position : \mathbb{R}^n,\ slope : [0, 1],\ line\_positions : \mathbb{R}^{n^*}\}; R_1, R_2, R_3)$, where R_1, R_2, and R_3 are defined as above.

Next, we must specify a Vocabulary $\Sigma = (\mathbf{C}, \mathbf{R}, \mathbf{V})$, consisting of a set of constant symbols $\mathbf{C}$, a set of relation symbols $\mathbf{R}$, and a set of variables $\mathbf{V}$ used as the signature for the state we are representing. We define $\mathbf{C}$ as $\{x, y, m, l_1, l_2\}$; $\mathbf{R}$ as the set containing $through(p, l)$, which holds when point p lies on line l; $s(p, x)$, which holds when photon p is observed at point x; $slope\_of\_one(l)$, which holds when the slope of line l is 1; and $\mathbf{V}$ as $\emptyset$, that is, the empty set.

Finally, we must provide an interpretation of the relation symbols of Σ into A. That is, as dictated by the formal machinery of (Arkoudas & Bringsjord 2009*b*), we must provide a mapping I that assigns to each relation symbol $R \in \mathbf{R}$ of arity n:

1. a relation $R^I \in \mathcal{R}$ of some arity m, called the **realization** of R:

$$R^I \subset A_{i_1} \times \cdots \times \subset A_{i_k}$$

   (where we might have $m \neq n$); and

2. a list of m pairs

$$[(l_{i_1}, j_1), ..., (l_{i_m}, j_m)],$$

   the **profile** of R, denoted by $Prof(R)$, with $1 \leq j_x \leq n$ for each $x = 1, ..., m$.

We summarize this interpretation in the following table.

Symbol	**Arity**	**Realization**	**Profile**
through	2	R_1	$[(position, 1), (line\_positions, 2)]$
observes	2	R_2	$[(position, 1), (position, 2)]$
slope_of_one	2	R_3	$[(slope, 1)]$

We now proceed by cases. Vivid has a control-construct built in for the formalization of proofs by cases.

> **Case I** (12): y is not on either of these two lines. In this case, assume that m observes a photon p at x, and observes the same photon p at y. However, the line connecting x and y would have a slope other than 1, which implies that the photon would be traveling at a speed other than the speed of light. This contradicts the light axiom. It follows that it is impossible that m should observe the same event at x and y.
>
> **Case II** (13): y is on one of these lines. Consider the other line $l1$, also of slope 1, through x. Let there be a photon q somewhere on this line other than at x. This photon may legitimately have traveled from x to its current location in space-time; hence q represents an event m observes at x. Now assume that m observes the same event at y. This would mean that q is somewhere on line $l2$, the line of slope 1 through y which is other than the line passing through x. Then, q must be on both lines $l1$ and $l2$, and hence at the point where they meet. However, there is no such point, since $l1$ and $l2$ are parallel to each other. Hence, m cannot observe the event represented by q at y. It follows that it is impossible that m should observe the same event at x and y. **QED**

We now define an additional axiom from our relations defined in Σ above, and proceed to demonstrate our proof within the framework of Vivid. We define the *not_on_same_worldline axiom* as the Horn clause: $observes(p, x) \wedge through(x, l) \wedge \neg through(y, l) \rightarrow \neg observes(p, y)$, taken to mean informally that if an observer observes a photon p at point x and x is on worldline l

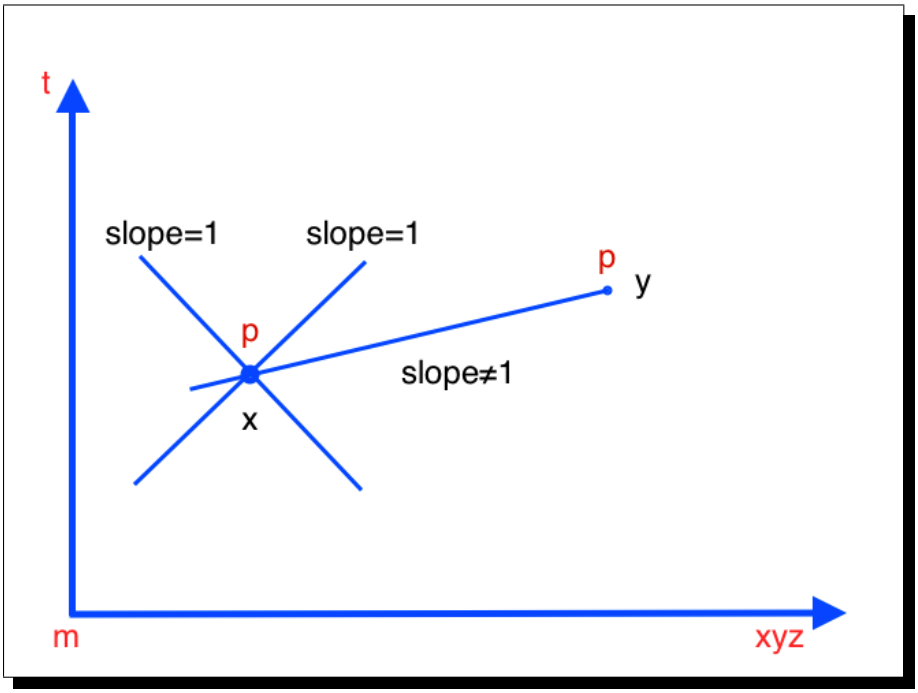


Figure 12: Case I

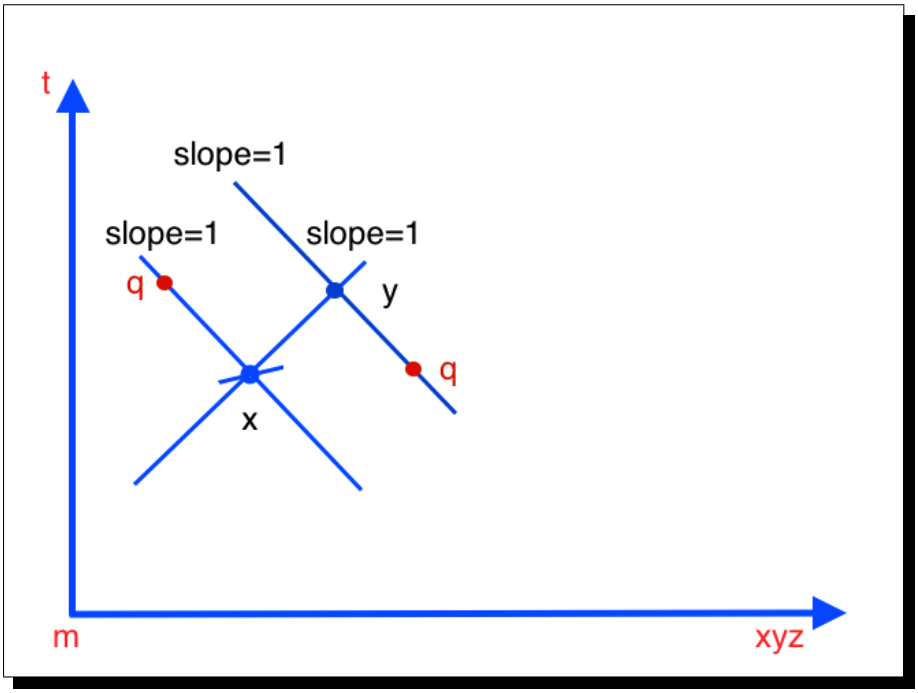


Figure 13: Case II

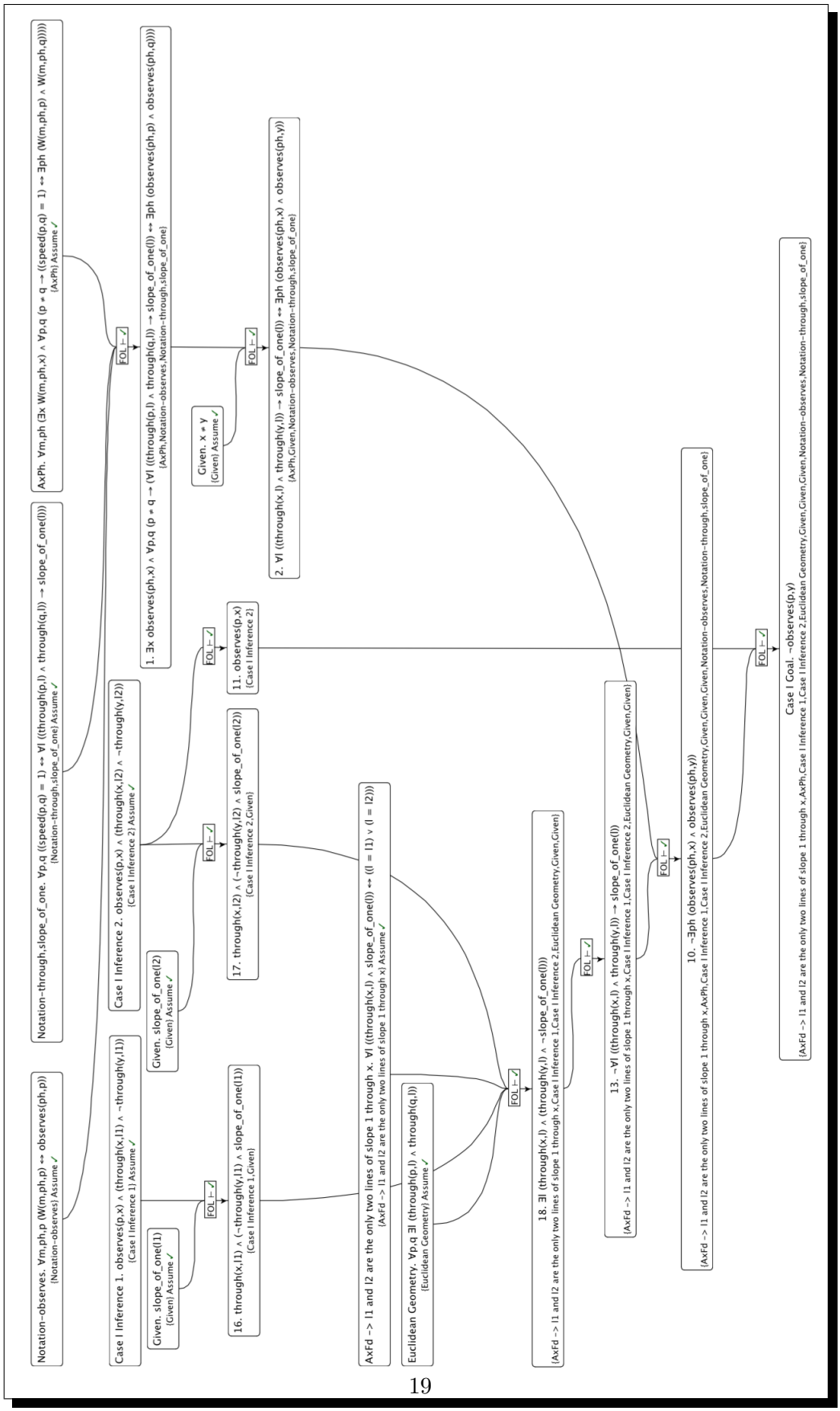


Figure 14: Proof for Case I

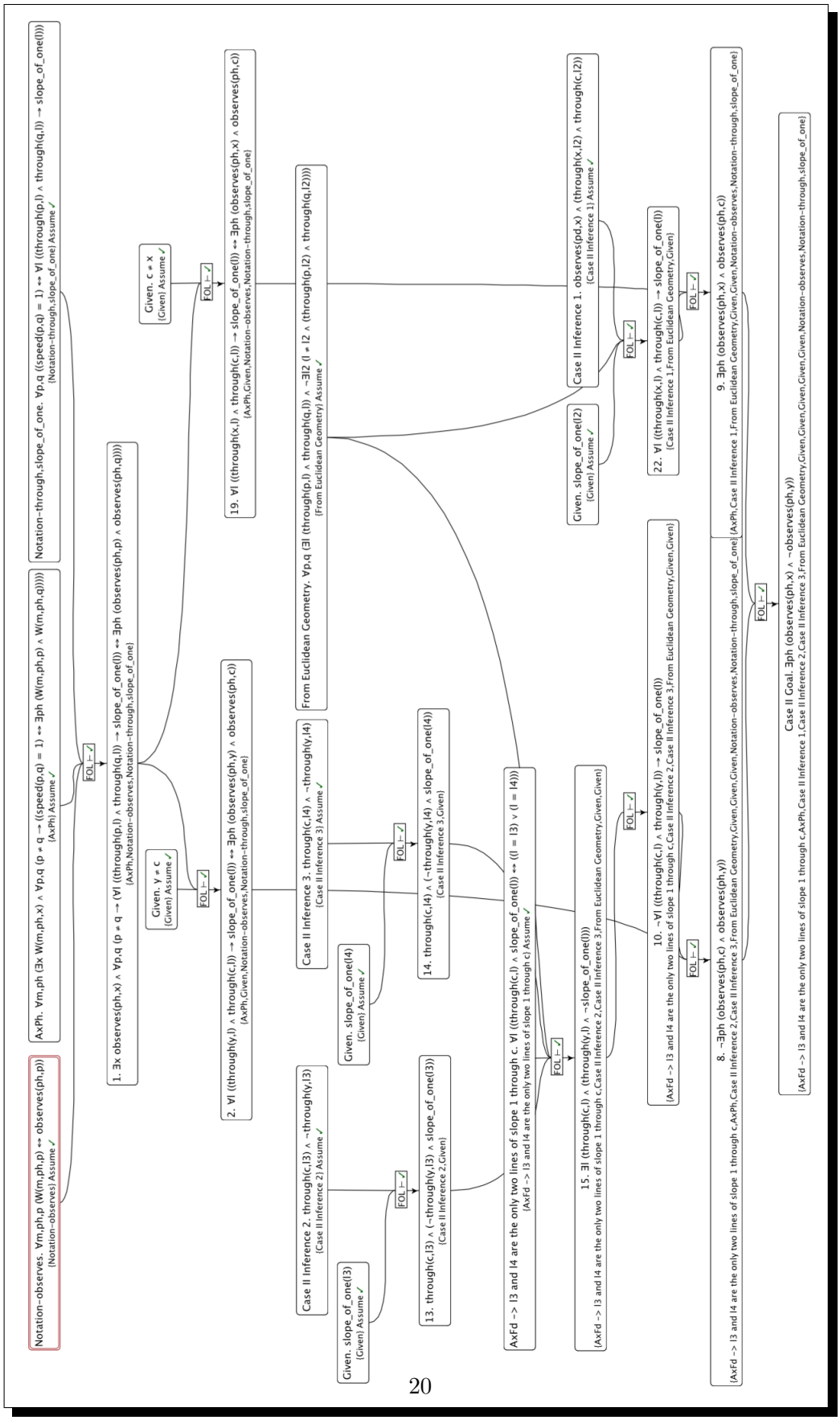


Figure 15: Proof for Case II

and y is not on worldline l, then that same photon p is not observable at y. Beginning with diagram Δ_0, we apply the diagrammatic rule $[C_1]$ with the field axiom $AxFd$ to derive two new diagrams, Δ_1 and Δ_2, corresponding to the cases above, respectively. Formally, we have, **cases** from $AxFd : (\sigma_1, p_1) \rightarrow \Delta_1 \mid (\sigma_2, p_2) \rightarrow \Delta_2$, where y is either on a worldline or is not. These cases are clearly exhaustive, which satisfies the side condition necessary to implement $[C_1]$.

> **Case I**: From diagram Δ_1, we claim a photon p exists at point x and use the diagrammatic rule $[Diagram - Reiteration]$ to derive Δ_3 . We proceed by using the **observe**[10] rule to extract sentential information from the diagram. Specifically, we extract two observations: **observe** $observes(p, x) \wedge through(x, l_1) \wedge \neg through(y, l_1)$, and **observe** $observes(p, x) \wedge through(x, l_2) \wedge \neg through(y, l_2)$. Through application of the *not_on_same_worldline axiom* with both of the observations obtained from Δ_3, we conclude $\neg observes(p, y)$; that is, m does not observe photon p at y, but does observe p at x. The semi-automated proof of this is formalized in Figure 14, in the hypergraph-based proof-engineering environment Slate. So, the set of bodies observed by m at x is different than the set of bodies m observes at y, completing the first case.
>
> **Case II**: In the case of Δ_2, x and y lie on the same worldline l_1 depicted in the diagram. However, $AxFd$ specifies another line of slope 1 through x. With this in mind, we apply the diagrammatic rule $[\Delta; \Delta]$ with $AxFd$ to derive diagram Δ_4. Now, claiming photon p' exists at point c on the new worldline l_2 that goes through x (where $c \neq x$), and utilizing the $[Diagram - Reiteration]$, we derive the new diagram Δ_5. We now **observe** $observes(p', x) \wedge through(x, l_2) \wedge \neg through(y, l_2)$. $observes(p', x)$ implies $observes(p', c)$ as x and c are both on l_2. However, $observes(p', c))$ implies $\neg observes(p', y)$ since $AxFd$ tells us that any line between c and y has slope $k \neq 1$. Then, photon p' has a speed different from the speed of light which is prohibited by the light axiom $AxPh$. Hence, m observes photon p' through x but not through y, the semi-automated proof of which is formalized in Figure 15, again in Slate. This gives us a body in the set of bodies that m observes at x but which m does not observe at y. **QED**

5.1.2 Enhancements to πVivid

The πVivid library, created during the 1-year extension, is a key part of the resources for achieving GCI; the library is written entirely in Python 2.7, is a concrete implementation of the Vivid framework capable of leveraging the unique inference schemata of Vivid (as invented and specified in Arkoudas & Bringsjord 2009*b*), as well as representing and dealing with incomplete information. Figure 16 gives an overall view of how πVivid performs heterogeneous natural deduction. The πVivid library was designed with ease of use, extensibility, and stability in mind. Along with an implementation of the Vivid framework, πVivid provides unit testing for every single operation that can be performed, ensuring that even after modifications to the source code, stability is ensured. In light of this, modifications to the source code are encouraged, and hence extensibility is enabled. In addition to vanilla evaluations of mathematical and logical statements during the semantic-value assignment process central to the Vivid framework (and thus to πVivid), we created a protocol that extends the πVivid library with new objects to be used during the course of deduction. This feature is especially handy when more complex expressions must be evaluated. Such expressions

[10] This rule is used to extract sentential information from diagrams. This is directly inspired by simple visual perception, and corresponds to the perception operator **P** in $\mathcal{DCEC}^*$.

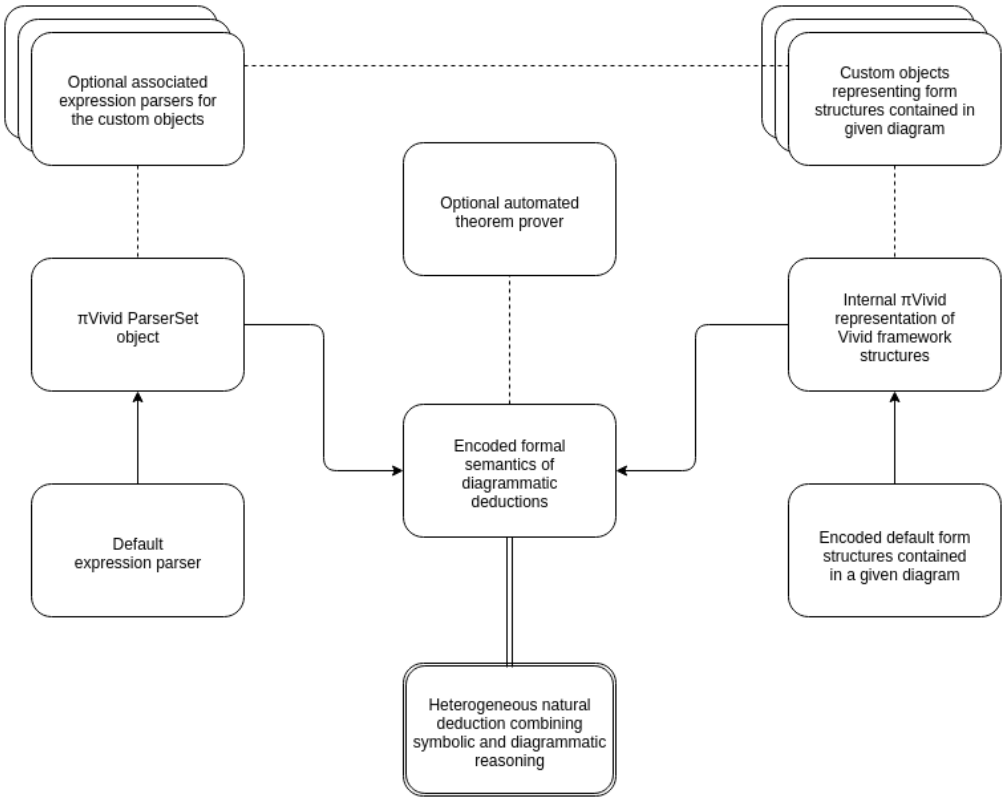


Figure 16: Overview of the Architecture of πVivid. Dashed lines represent optional behavior flowing between components, while solid lines represent necessary behavior flowing between components in the direction of the arrowhead.

can be evaluated both easily and concisely; additionally, πVivid also provides mechanisms for linking to an automated theorem prover if further reasoning must be done before the evaluation of a particular statement. The πVivid library also provides a convenient mechanism for the detection of incorrect inference steps through the **observe** semantic form; we have recently used this feature, in the context of automated tutoring, to alert a student that an incorrect step had been made during the construction of the proof of Theorem 2.2 from (Andréka et al. 2007).[11] With the introduction of the πVivid library, the functionality of the Vivid framework is now freely available to programmers and logicians alike, indeed to the entire AI community.

The πVivid library remains under continuous development. In the latest installment under the 1-year extension, support for importing and exporting purely symbolic expressions was added. This new feature has a profound impact on the πVivid library's ability to carry out heterogeneous proofs. Now, purely symbolic expressions can be exported and reasoned over as premises and conclusions by an automated theorem prover, and entirely new sentences derived from the exported sentences can be imported back into πVivid and used by the diagrammatic sentential forms contained therein, seamlessly. With this feature, the domain of proofs verifiable by πVivid has been extended greatly,

[11] A full video of this achievement, with prominent attribution to AFOSR, is available here: http://rair.cogsci.rpi.edu/projects/artificial-intelligence-in-physics-and-mathematics-education

and the feature itself is incredibly easy to use, requiring only a single function call for both importing and exporting, respectively.

5.2 In (Early) Automated Discovery in Formal Biology

Over and over again, we have encountered reviews and discussions underscoring an express need for formalism in biology (Lux, Bramlett, Ball & Peccoud 2012). In this context, of note is the fact that logicist formal techniques have numerous virtues (not least of them that such techniques originally gave rise to computer science and AI), and now, via the GCI paradigm, they place 50 years of AI research at biology's service. In physics and other fields, artificial discovery has yielded useful and previously unknown results; witness our own work. In biology, the stage may not be quite yet fully set, but our survey of BioCAD in particular leads us to believe that it's goals are remarkably aligned with our own in the GCI paradigm, though its methods so far have largely been different.

In general terms, we were (and continue to be) interested in applying our GCI resources to artificial, automated, or semi-automated scientific discovery in biology. Research on computational scientific discovery aims to develop computer systems which produce results that, if a human scientist did the same, we would refer to as 'advances' in science (this is in line, recall, with §2). We have accomplished this for aspects of mathematical logic and formal physics; and we took steps toward erecting a full foundation and mature springboard for machine discovery in these disciplines. We also started to do the same for a notoriously informal empirical science: biology.

5.2.1 Prior Work and Some of Our Accomplishments

5.2.1.1 Woodger, & Our Initial Mechanization Thereof

Toward the end of Phase 1 and before our 1-year extension, we discovered the work of the remarkable biologist J. H. Woodger, who, in (Woodger 1937), proposed "axioms" of the Mendelian genetics of diploid eukaryotes. Heavily influenced by Russell and Whitehead's *Principia Mathematica*,[12] and the movement of logical positivism, Woodger attempted to develop the Mendelian theory by deducing it's "theorems" from his axioms. *Principia* attempted this for mathematics, and the fascinating historical sub-plot, which of course we leave aside here, is that none other than Herbert Simon, at the dawn of modern AI at the famous 1956 Dartmouth conference, stole the show by arriving on scene with a computer program, LOGIC THEORIST, that could automatically prove some substantive theorems from none other than *Principia*.

From scientific, engineering, and historical standpoints, Woodger's choice of formal, extensional logic as a language suggests the possibility of exploiting automated theorem proving within his framework — an option Woodger did not have in 1937 (two more decades had to pass before Simon founded automated theorem proving at the Dartmouth conference). The RAIR Lab specializes in formal modelling and automated theorem proving, central of course to the GCI paradigm and program, and RAIR-Lab researcher A. Sen obtained machine-discovered proofs of several of Woodger's theorems. These are the first such formal, and machine-verified, proofs, to our knowledge. One of

[12] The three-volume set is available online and free in the Univ. of Michigan's Historical Math Collection:

Vol I http://quod.lib.umich.edu/cgi/t/text/text-idx?c=umhistmath;idno=AAT3201.0001.001

Vol II http://quod.lib.umich.edu/cgi/t/text/text-idx?c=umhistmath;idno=AAT3201.0002.001

Vol III http://quod.lib.umich.edu/cgi/t/text/text-idx?c=umhistmath;idno=AAT3201.0003.001

them was presented briefly at our team's review at the conclusion of Phase 1 of the GCI research program, in Arlington; during our current 1-year extension, we made further progress (see below).

Woodger provides manual, paper-and-pencil proofs of several of his theorems in an appendix to his book; the first theorem is 1.3.6. An automated proof of this in Bringsjord's Slate system is shown in Figure 17. Woodger's axiomatic system is founded on Alfred Tarski's formulation of the axioms of **mereology**, the theory of parthood relations. Mereology, like Woodger's axiomatic biology, has never received serious computational treatment. Figure 18 shows the automated proofs, in the Slate system, of the first theorems of Tarskian mereology. Further, Figure 19 shows a semi-automated proof of one of these theorems, which corresponds closely to a manual proof obtained by Sen. These results provide a perfect foundation from which to move further into the "GCI-ification" of Woodger's work.

Figure 17: Modernized Version of Very First Theorem Woodger Proves

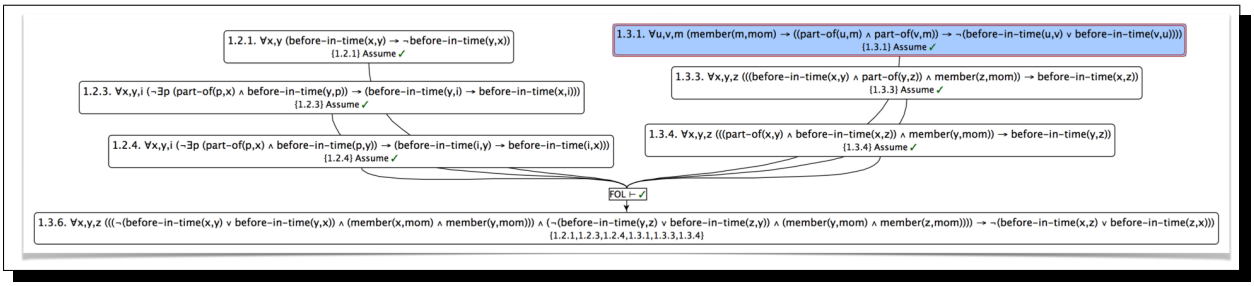
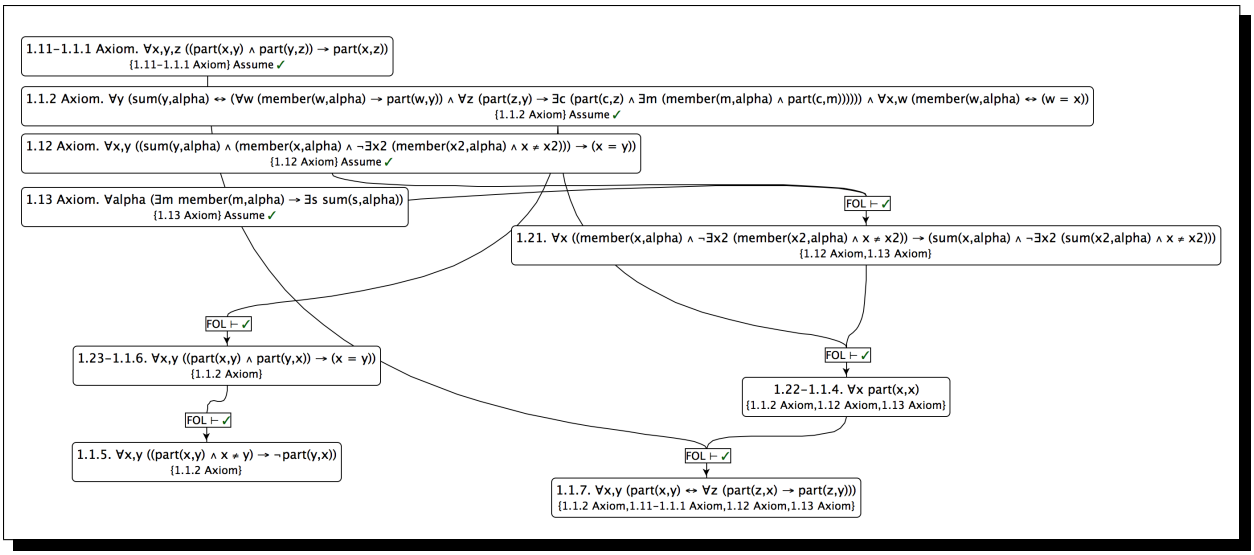
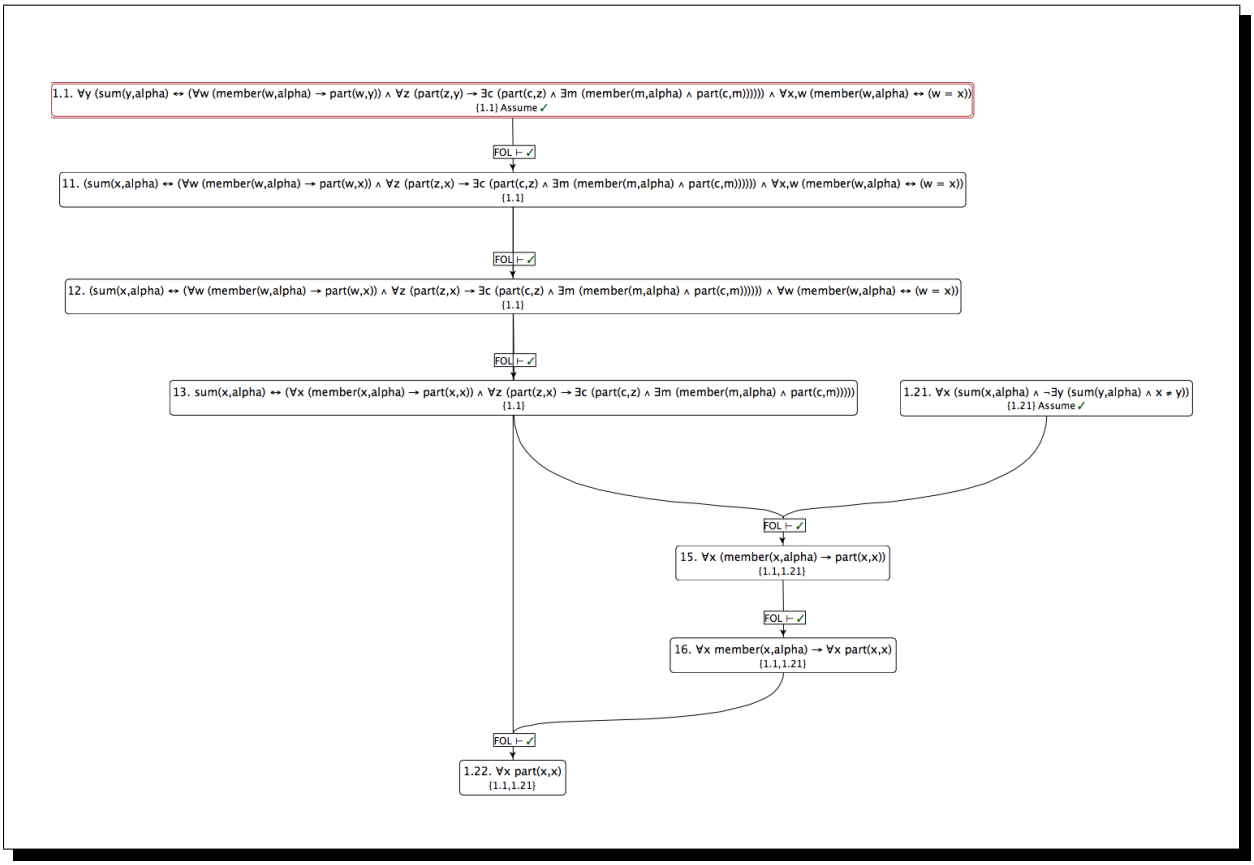


Figure 18: The Initial Development of Mereology



During our 1-year extension, we made further progress: we obtained proofs of several theorems concerning Woodger's mereological structure (with some temporal information) of the Mendelian theory of diploid eukaryotes. In section 1.2 of Woodger's (1937) *The Axiomatic Method in Biology*, he further expounds the mereological foundations of his formal biological theory. Using the

Figure 19: Semi-Automated Proof of Theorem 1.22

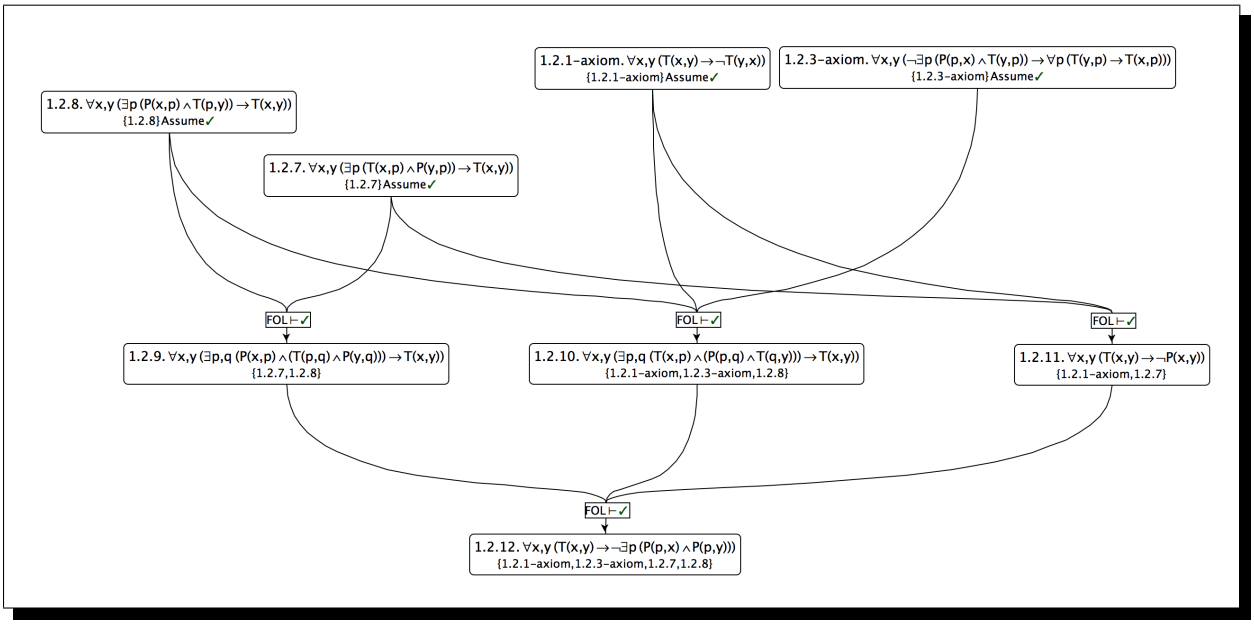


Woodger-Tarski notion of parthood and temporal precedence, the proof of Theorem 1.2.9 establishes that, for all individual objects x and y, the existence of objects p and q such that x is a part of p, p precedes q in time, and y is a part of q, implies that x precedes y in time. The proof of Theorem 1.2.11 establishes that if x precedes y in time, then x cannot be a part of y, and the proof of Theorem 1.2.12 establishes that if x precedes y in time, there does not exist a common part between them. The proofs of Theorems 1.2.9–1.1.12 were obtained fully automatically by use of the theorem prover SNARK, a prover callable from within Slate (as is ShadowProver, which should be able to speed up the automated proof-discovery results here). Figure 20 shows a depiction of the requirements for proving each theorem (i.e., a proof plan), in Slate.

5.3 In Analogico-*In*ductive Reasoning (AIR)

As we have made clear, our r&d hitherto has focused on "great computational intelligence" (GCI) based on analogico-*de*ductive reasoning (ADR); hence we have referred to 'GCI-ADR.' What, then, about the marriage of analogical processing with not deduction (or not *only* with deduction), but rather with induction, that is, with forms of rigorous argument that don't reach a level of proof, but are nonetheless still sufficiently compelling to move science along. (These arguments are often probabilistic in nature.) While the *formal* sciences — which has been a large part of our domain

Figure 20: Newly Obtained Proofs from (Woodger 1937, §1.2)



in work carried out so far — put a premium on proof, certainly the *empirical* sciences, overall, do not. If we are going to be able to extend our general approach of GCI from the formal sciences to the empirical sciences, and to real-world adversarial domains, we are clearly going to need to extend from GCI-A*D*R to GCI on the basis of analogico-*in*ductive reasoning (A*I*IR).

The first step in it is to invent some new inductive logics that are able to be hooked to automated reasoning systems of the sort that we have leveraged in the project so far (e.g., the SNARK automated theorem prover, used for our work in GCI in the axiomatic-physics realm; see Stickel, Waldinger, Lowry, Pressburger & Underwood 1994). To our knowledge, no one has done this before. In the current landscape, instead, the computational harnessing of traditional mathematical statistics (as e.g. covered in Hogg, Craig & McKean 2005) is used to power statistical learning techniques that by definition fail to provide justification in the classical forms used by scientists to argue and counter-argue about the issues at hand.

5.3.1 Accomplishments in the 1-Year Extension

As a first step in our exploration of analogico-inductive reasoning (AIR), we use $\mathcal{CEC}$ to formalize argument schemas commonly used in everyday analogical reasoning. Although argument by analogy is studied and featured in many computational models, less appreciated is the ability to reason *over* analogies (RoA); that is, not only being able to produce inferences in accordance with arguments by analogy, but having the ability to negate analogies, recognize and learn to avoid bad analogies, compare the relative strengths of analogies, reason about them nonmonotonically, evaluate hypothetical analogies, and so on. To do all of these things, one needs the ability to represent analogies (and not just the products of analogies) in such a way that the analogies themselves can be objects of reasoning processes (including analogy). To our knowledge, no one has imbued a computing machine with this ability. During our 1-year extension, we took a first step toward the

full ability, on the part of a computational agent, to reason over analogies through a formalization, based on $\mathcal{CEC}$, that treats analogical mappings and hypothetical inferences as objects between which confidence can be propagated.

For example, consider the following argued to a teenager (we label this informal example $\mathbf{NL}_f$):

> If work is like college, and working hard at work means I get more money, then working hard at school means I get a better education. In fact, working hard at work does yield more money, and work is indeed like college (with respect to money and education). Therefore, working hard at school yields a better education.

We argue in (Licato & Fowler forthcoming) that example $\mathbf{NL}_f$ can be captured formally by the following schema:

$$\frac{\mathbf{B}^x(\varphi_S), \mathbf{B}^y(M_{ST}), \mathbf{B}^w((M_{ST} \wedge \varphi_S) \rightarrow \varphi_T)}{\mathbf{B}^{\min(w,x,y)}(\varphi_T)} f \tag{2}$$

The notation used here is based on the cognitive event calculus ($\mathcal{CEC}$) (Arkoudas & Bringsjord 2009*a*, Bringsjord, Govindarajulu, Licato, Sen, Johnson, Bringsjord & Taylor 2015), though here the agent and time parameters are omitted for simplicity. Note that the notation for strength factors (denoted here as superscripts following the modal operator $\mathbf{B}$) is a new addition. The strength factor of the conclusion is a minimum of w, x, and y, using the assumption sometimes referred to as the *weakest link principle* (Pollock 1995).

For some source S and target domain T, we use φ_S and φ_T to denote formulae belonging to those respective domains. M_{ST} is a shorthand for the formula stating that: (1) M is a mapping between domains S and T, (2) φ_S and φ_T are formulae in S and T, respectively, and (3) φ_S and φ_T are mapped to each other in M.

The informal example $\mathbf{NL}_f$ and its formal schema correspond to what is often called *argument-by-analogy*. But human-level thought contains many analogical arguments of different forms, for instance the argument implied in informal example $\mathbf{NL}_{b1}$. Imagine a teenager presented with $\mathbf{NL}_f$ and responding with:

> If work is like college, and working hard at work means I get more money, then working hard at school means I get a better education. I accept that. But I worked hard at school and all I got was the ability to pass tests; I did not get a better education. Meanwhile, working hard at work does in fact yield better pay. Therefore, work is not like college.

Such an argument seems to conclude that there should be a reduction in the confidence of an analogical mapping, unlike in Schema 2. We capture $\mathbf{NL}_{b1}$ with the following schema:

$$\frac{\mathbf{B}^x(\varphi_S), \mathbf{B}^y(\neg\varphi_T), \mathbf{B}^w((M_{ST} \wedge \varphi_S) \rightarrow \varphi_T)}{\mathbf{B}^{\min(w,x,y)}(\neg M_{ST})} b1 \tag{3}$$

Our approach to formalizing analogies and analogical mappings themselves as objects in $\mathcal{CEC}$ has already shown a nice theoretical result. We can show (and this is proven formally in Licato and Fowler (2016)) that certain arguments by analogy cannot have the same weight as those based on Schemas 2 and 3 (under certain assumptions of deductive strength). For example, the informal argument $\mathbf{NL}_x$:

If work is like college, and working hard at work means I get more money, then working hard at school means I get a better education. I accept that. But work is simply not like college; it's like comparing apples to oranges. Therefore, working hard at school won't lead to a better education.

This corresponds to a schema that cannot have the same validity as Schemas 2 and 3. Licato and Fowler (2016) also show that under certain assumptions, we can list all valid forms of argument by analogy; these are pictured in Figure 21.

Warrant	$\mathbf{B}^{w}((M_{ST} \wedge \varphi_S) \rightarrow \varphi_T)$		
Premise 1	$\mathbf{B}^{x}(\varphi_S)$	$\mathbf{B}^{x}(\neg\varphi_T)$	
Premise 2	$\mathbf{B}^{y}(M_{ST})$	$\mathbf{B}^{y}(M_{ST})$	$\mathbf{B}^{y}(\varphi_S)$
Conclusion	$\mathbf{B}^{\min(w,x,y)}(\varphi_T)$	$\mathbf{B}^{\min(w,x,y)}(\neg\varphi_S)$	$\mathbf{B}^{\min(w,x,y)}(\neg M_{ST})$
Inference Name	f	b1	b2

Warrant	$\mathbf{B}^{w}((\varphi_S \wedge \varphi_T) \rightarrow M_{ST})$		
Premise 1	$\mathbf{B}^{x}(\varphi_S)$	$\mathbf{B}^{x}(\neg M_{ST})$	
Premise 2	$\mathbf{B}^{y}(\varphi_T)$	$\mathbf{B}^{y}(\varphi_S)$	$\mathbf{B}^{y}(\varphi_T)$
Conclusion	$\mathbf{B}^{\min(w,x,y)}(M_{ST})$	$\mathbf{B}^{\min(w,x,y)}(\neg\varphi_T)$	$\mathbf{B}^{\min(w,x,y)}(\neg\varphi_S)$
Inference Name	f	b1	b2

Figure 21: Valid Analogical Argument Forms for Two Different Warrants, $\mathcal{W}$ (top table) and $\mathcal{W}'$ (bottom table)

References

Andréka, H., Madarász, J. X. & Németi, I. (2007), Logic of Space-Time and Relativity Theory, *in* M. Aiello, I. Pratt-Hartmann & J. V. Benthem, eds, 'Handbook of Spatial Logics', Springer, pp. 607–711.

Arkoudas, K. (2000), Denotational Proof Languages, PhD thesis, Massachusetts Institute of Technology.

Arkoudas, K. & Bringsjord, S. (2009*a*), 'Propositional Attitudes and Causation', *International Journal of Software and Informatics* **3**(1), 47–65.
URL: *http://kryten.mm.rpi.edu/PRICAI_w_sequentcalc_041709.pdf*

Arkoudas, K. & Bringsjord, S. (2009*b*), 'Vivid: An AI Framework for Heterogeneous Problem Solving', *Artificial Intelligence* **173**(15), 1367–1405. The url http://kryten.mm.rpi.edu/vivid/vivid.pdf provides a preprint of the penultimate draft only. If for some reason it is not working, please contact either author directly by email.
URL: *http://kryten.mm.rpi.edu/vivid_030205.pdf*

Bringsjord, S. (1998), 'Chess is Too Easy', *Technology Review* **101**(2), 23–28.
URL: *http://kryten.mm.rpi.edu/SELPAP/CHESSEASY/chessistooeasy.pdf*

Bringsjord, S., Govindarajulu, N., Ellis, S., McCarty, E. & Licato, J. (2014), 'Nuclear Deterrence and the Logic of Deliberative Mindreading', *Cognitive Systems Research* **28**, 20–43.
URL: *http://kryten.mm.rpi.edu/SB_NSG_SE_EM_JL_nuclear_mindreading_062313.pdf*

Bringsjord, S., Govindarajulu, N. S., Licato, J., Sen, A., Johnson, J., Bringsjord, A. & Taylor, J. (2015), On Logicist Agent-Based Economics, *in* 'Proceedings of Artificial Economics 2015 (AE 2015)', University of Porto, Porto, Portugal.

Bringsjord, S. & Licato, J. (2012), Psychometric Artificial General Intelligence: The Piaget-MacGuyver Room, *in* P. Wang & B. Goertzel, eds, 'Foundations of Artificial General Intelligence', Atlantis Press, Amsterdam, The Netherlands, pp. 25–47. This url is to a preprint only.
URL: *http://kryten.mm.rpi.edu/Bringsjord_Licato_PAGI_071512.pdf*

Bringsjord, S., Taylor, J., Shilliday, A., Clark, M. & Arkoudas, K. (2008), Slate: An Argument-Centered Intelligent Assistant to Human Reasoners, *in* F. Grasso, N. Green, R. Kibble & C. Reed, eds, 'Proceedings of the 8th International Workshop on Computational Models of Natural Argument (CMNA 8)', University of Patras, Patras, Greece, pp. 1–10.
URL: *http://kryten.mm.rpi.edu/Bringsjord_etal_Slate_c mna_crc_061708.pdf*

Govindarajalulu, N. S., Bringsjord, S. & Taylor, J. (2015), 'Proof Verification and Proof Discovery for Relativity', *Synthese* **192**(7), 2077–2094.

Govindarajulu, N., Licato, J. & Bringsjord, S. (2013), Small Steps Toward Hypercomputation via Infinitary Machine Proof Verification and Proof Generation, *in* M. Giancarlo, A. Dennuzio, L. Manzoni & A. Porreca, eds, 'Unconventional Computation and Natural Computation; Lecture Notes in Computer Science; Volume 7956', Springer-Verlag, Berlin, Germany, pp. 102–112.

Govindarajulu, N., Licato, J. & Bringsjord, S. (2014), Toward a Formalization of QA Problem Classes, *in* B. Goertzel, L. Orseau & J. Snaider, eds, 'Artificial General Intelligence; LNAI 8598', Springer, Switzerland, pp. 228–233.
URL: *http://kryten.mm.rpi.edu/NSG_SB_JL_QA_formalization_060214.pdf*

Govindarajulu, N. S. & Bringsjord, S. (2014), 'Proof Verification Can be Hard!'. Presented at the 10th Conference of Computability in Europe (CiE).
URL: *http://www.naveensundarg.com/papers/ProofVerificationCanBeHard.pdf*

Hogg, R., Craig, A. & McKean, J. (2005), *Introduction to Mathematical Statistics (6th ed)*, Pearson Prentice Hall, New Delhi, India.

Holyoak, K. J., Gentner, D. & Kokinov, B. N. (2001), Introduction: The Place of Analogy in Cognition, *in* D. Gentner, K. J. Holyoak & B. N. Kokinov, eds, 'The Analogical Mind: Perspectives from Cognitive Science', The MIT Press, chapter 1.

Hummel, J. E. & Holyoak, K. J. (1997), 'Distributed Representations of Structure: A Theory of Analogical Access and Mapping', *Psychological Review* **104**(3), 427–466.

Hummel, J. E. & Holyoak, K. J. (2003), 'Relational Reasoning in a Neurally-plausible Cognitive Architecture: An Overview of the LISA Project', *Cognitive Studies: Bulletin of the Japanese Cognitive Science Society* **10**, 58–75.

Jaśkowski, S. (1934), 'On the Rules of Suppositions in Formal Logic', *Studia Logica* **1**, 5–32.

Kurzweil, R. (2006), *The Singularity Is Near: When Humans Transcend Biology*, Penguin USA, New York, NY.

Licato, J. & Bringsjord, S. (2015), PAGI World: A Simulation Environment to Challenge Cognitive Architectures, *in* 'IBM Cognitive Systems Institute Group Speaker Series'.

Licato, J., Bringsjord, S. & Hummel, J. E. (2012), Exploring the Role of Analogico-Deductive Reasoning in the Balance-Beam Task, *in* 'Rethinking Cognitive Development: Proceedings of the 42nd Annual Meeting of the Jean Piaget Society', Toronto, Canada.

Licato, J. & Fowler, M. (forthcoming), Formalizing Confidence Propagation in Analogico-Inductive Reasoning, *in* 'Proceedings of IACAP 2016'.

Licato, J., Govindarajulu, N. S., Bringsjord, S., Pomeranz, M. & Gittelson, L. (2013), Analogico-Deductive Generation of Gödel's First Incompleteness Theorem from the Liar Paradox, *in* F. Rossi, ed., 'Proceedings of the 23rd International Joint Conference on Artificial Intelligence (IJCAI–13)', Morgan Kaufmann, Beijing, China, pp. 1004–1009. Proceedings are available online at `http://ijcai.org/papers13/contents.php`. The direct URL provided below is to a preprint. The published version is available at `http://ijcai.org/papers13/Papers/IJCAI13-153.pdf`.
URL: *http://kryten.mm.rpi.edu/ADR_2_GI_from_LP.pdf*

Licato, J., Marton, N., Dong, B., Sun, R. & Bringsjord, S. (2015), Modeling the Creation and Development of Cause-Effect Pairs for Explanation Generation in a Cognitive Architecture, *in* 'Proceedings of the 2015 International Workshop on Artificial Intelligence and Cognition (AIC 2015)'.

Lux, M. W., Bramlett, B. W., Ball, D. A. & Peccoud, J. (2012), 'Genetic Design Automation: Engineering Fantasy or Scientific Renewal?', *Trends in Biotechnology* **30**(2), 120–126.

Manzano, M. (1996), *Extensions of First Order Logic*, Cambridge University Press, Cambridge, UK.

Pollock, J. L. (1995), *Cognitive Carpentry: A Blueprint for How to Build a Person*, MIT Press.

Russell, S. & Norvig, P. (2009), *Artificial Intelligence: A Modern Approach*, Prentice Hall, Upper Saddle River, NJ. Third edition.

Stannett, M. & Németi, I. (2014), 'Using Isabelle/HOL to Verify First-Order Relativity Theory', *Journal of Automated Reasoning* **52**(4), 361–378.

Stickel, M., Waldinger, R., Lowry, M., Pressburger, T. & Underwood, I. (1994), Deductive Composition of Astronomical Software From Subroutine Libraries, *in* 'Proceedings of the Twelfth International Conference on Automated Deduction (CADE–12)', Nancy, France, pp. 341–355. SNARK can be obtained at the url provided here.
URL: *http://www.ai.sri.com/∼stickel/snark.html*

Wang, H. (1995), On 'Computabilism' and Physicalism: Some Subproblems, *in* J. Cornwell, ed., 'Nature's Imagination: The Frontiers of Scientific Vision', Oxford University Press.

Woodger, J. (1937), *The Axiomatic Method in Biology*, Cambridge University Press, Cambridge, UK.